Fred Bunz

Principles of Cancer Genetics

Springer

Principles of Cancer Genetics

Fred Bunz

Principles of Cancer Genetics

Springer

Fred Bunz, MD, PhD
Johns Hopkins University School of Medicine
Baltimore, Maryland
USA

Cover illustration: Fluorescence-labeled DNAs migrate through the capillaries of an automated DNA sequencing apparatus. Each fluorescent spot represents a distinct nucleotide. Image courtesy of Devin Dressman, PhD, Johns Hopkins University.

ISBN 978-1-4020-6783-9 e-ISBN 978-1-4020-6784-6

Library of Congress Control Number: 2007938449

© 2008 Springer Science + Business Media B.V.
No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Preface

Nothing makes sense in biology, but in the light of evolution.

Theodosius Dobzhansky (1900–1975)

Cancer is caused by genetic alterations. To understand the nature of these alterations – how they arise and how they are inherited – is to grasp the essence of cancer. In the past several years, hundreds of genes have been categorized as cancer genes. These genes, in turn, have illuminated basic pathways and regulatory networks that control cell fate. The identification of cancer genes and their respective functions in cells and tissues has revolutionized our view of tumors and how they grow. The primary literature that describes these intellectual strides is daunting in scope, but the central ideas can be readily condensed and simplified. The intent behind this book is to provide context to recent advances in cancer research by outlining basic principles that describe how cancer genes arise, are inherited, and function. Although the list of recognized cancer genes is likely to grow rapidly in the coming years, the fundamental principles of cancer genetics will likely endure.

This book is aimed at advanced undergraduates who have completed introductory courses in genetics, biology and biochemistry, and at medical students. There are several excellent texts that provide an overview of cancer biology and genetics, including *The Biology of Cancer* by Weinberg and *The Genetic Basis of Human Cancer* by Vogelstein and Kinzler. In contrast to these comprehensive texts, this modest book is focused on the most highly representative genes that underlie the most common cancers. Attention is primarily devoted to cancer genes and the application of evolutionary theory to explain why the cell clones that harbor cancer genes tend to expand. Areas of controversy are avoided, in favor of firmly established concepts. This book does not delve into tumor pathobiology beyond what is required to understand the role of genetic alterations in neoplastic growth. For students with a general interest in cancer, this book will provide an accessible overview. For students contemplating future study in the fields of oncology or cancer research, this book will be suitable as a primer. *Principles of Cancer Genetics* is intended not to replace existing texts but to complement them.

I am indebted to my teachers. The mentors I have been lucky to encounter have taught largely by example. Sanford Simon generously provided me with my first

undergraduate laboratory experience. Bruce Stillman, the supervisor of my doctoral research, introduced me to molecular biology and biochemistry as tools for rigorous cancer research. More recently, Bert Vogelstein and his partner Ken Kinzler have provided a model of incisive thinking, dedication, fearlessness, generosity and friendship that everyone should attempt to emulate. I am also indebted to my students, who challenge me in every way and fuel me with their energy and determination.

A career in science is filled with ups and downs. I have been lucky to have company on this journey. My girlfriend Karla Jusczyk, my friends and my family have lovingly supported me and kept me happily distracted. To all of these people I will be forever grateful.

Baltimore Fred Bunz
August 2007

Contents

Preface . . . . . **v**

1 The Genetic Basis of Cancer **1**

The Cancer Gene Theory 1
Cancers are Invasive Tumors 2
Cancer is a Unique Type of Genetic Disease 3
What are Cancer Genes and How are They Acquired? 4
Mutations Alter the Human Genome 5
Genes and Mutations 7
Genetic Variation and Cancer Genes 9
Which Mutations are Important in Cancer? 12
Single Nucleotide Substitutions 15
Gene Silencing by Cytosine Methylation: Epigenetics 18
Environmental Mutagens, Mutations and Cancer 18
Inflammation Promotes the Propagation of Cancer Genes 23
Darwinian Selection and the Clonal Evolution of Cancers 27
Selective Pressure and Adaptation: Hypoxia and Altered Metabolism . . . 29
Multiple Somatic Mutations Punctuate Clonal Evolution 30
How Many Mutations Contribute to a Cancer? 31
Colorectal Cancer: A Model for Understanding the Process of Tumorigenesis 35
Do Cancer Cells Divide More Rapidly than Normal Cells? 40
Germline Cancer Genes Allow Neoplasia to Bypass Steps in Clonal Evolution 41
Cancer Syndromes Reveal Rate-limiting Steps in Tumorigenesis. 43
Understanding Cancer Genetics 45

2 Oncogenes **49**

What is an Oncogene? 49
The Discovery of Transmissible Cancer Genes 49
Viral Oncogenes are Derived from the Host Genome 52
The Search for Activated Oncogenes: The RAS Gene Family 54

Complex Genomic Rearrangements: The MYC Gene Family 57
Proto-oncogene Activation by Gene Amplification 58
Proto-oncogene Activation by Chromosomal Translocation 61
Chromosomal Translocations in Liquid and Solid Tumors 62
Chronic Myeloid Leukemia and the Philadelphia Chromosome 63
Ewing's Sarcoma and the Oncogenic Activation of a Transcription Factor 67
Oncogene Discovery in the Genomic Era: Mutations in *PIK3CA* 69
Selection of Tumor-Associated Mutations. 70
Multiple Modes of Proto-oncogene Activation 71
Oncogenes are Dominant Cancer Genes 72
Germline Mutations in *RET* and *MET* Confer Cancer Predisposition 73
Proto-oncogene Activation and Tumorigenesis 74

3 Tumor Suppressor Genes 77

What is a Tumor Suppressor Gene?. 77
The Discovery of Recessive Cancer Phenotypes 77
Retinoblastoma and Knudson's Two-Hit Hypothesis 79
Chromosomal Localization of the Retinoblastoma Gene 80
The Mapping and Cloning of the Retinoblastoma Gene 84
Tumor Suppressor Gene Inactivation: The Second 'Hit' and Loss of Heterozygosity 86
Recessive Genes, Dominant Traits. 87
APC Inactivation in Inherited and Sporadic Colorectal Cancers. 88
P53 Inactivation: A Frequent Event in Tumorigenesis 91
Functional Inactivation of p53: Tumor Suppressor Genes and Oncogenes Interact 93
Germline Inheritance of Mutant *P53*: Li–Fraumeni Syndrome 94
Cancer Predisposition: Allelic Penetrance, Relative Risk and Odds Ratios 98
Breast Cancer Susceptibility: *BRCA1* and *BRCA2* 101
Genetic Losses on Chromosome 9: *CDKN2A* 104
Complexity at *CDKN2A*: Neighboring and Overlapping Genes 106
Genetic Losses on Chromosome 10: *PTEN* 108
SMAD4 and the Maintenance of Stromal Architecture 111
Two Distinct Genes Underlie Neurofibromatosis 113
Multiple Endocrine Neoplasia Type 1 116
Most Tumor Suppressor Genes are Tissue-Specific 116
Modeling Cancer Syndromes in Mice 117
Tumor Suppressor Gene Inactivation During Colorectal Tumorigenesis . . 120
Inherited Tumor Suppressor Gene Mutations: Gatekeepers and Landscapers 122
Maintaining the Genome: Caretakers 123

4 Genetic Instability and Cancer 125

What is Genetic Instability? 125
The Majority of Cancer Cells are Aneuploid 126
Aneuploid Cancer Cells Exhibit Chromosome Instability 128
Chromosome Instability Arises Early in Colorectal Tumorigenesis 130
Chromosomal Instability Accelerates Clonal Evolution 131
What Causes Aneuploidy? 133
Transition from Tetraploidy to Aneuploidy During Tumorigenesis 135
Multiple Forms of Genetic Instability in Cancer 137
Defects in Mismatch Repair Cause Hereditary Nonpolyposis Colorectal Cancer 139
Mismatch Repair-Deficient Cancers Have a Distinct Spectrum of Mutations 145
Defects in Nucleotide Excision Repair Cause Xeroderma Pigmentosum 146
NER Syndromes: Clinical Heterogeneity and Pleiotropy 153
DNA Repair Defects and Mutagens Define Two Steps Towards Genetic Instability 154
Defects in DNA Crosslink Repair Cause Fanconi Anemia 156
A Defect in DNA Double-Strand Break Responses Causes Ataxia-telangiectasia 160
Bloom Syndrome Features Hyper-recombination 163
Aging and Cancer: Insights from the Progeroid Syndromes 166
Overview: Genes and Genetic Instability 170

5 Cancer Gene Pathways 173

What are Cancer Gene Pathways? 173
Cellular Pathways are Defined by Protein–Protein Interactions 174
Individual Biochemical Reactions, Multistep Pathways, and Networks . . 177
Protein Phosphorylation is a Common Regulatory Mechanism 180
Signals from the Cell Surface: Protein Tyrosine Kinases 181
Membrane-Associated GTPases: The RAS Pathway 186
Genetic Alterations of the RAS Pathway in Cancer 189
Membrane-Associated Lipid Phosphorylation: The PI3K/AKT Pathway 190
Genetic Alterations of the PI3K/AKT Pathway in Cancer 193
Morphogenesis and Cancer: The WNT/APC Pathway 194
Inactivation of the WNT/APC Pathway in Cancers 196
TGF-β/SMAD Signaling Maintains Tissue Homeostasis 198
C-MYC is a Downstream Effector of Multiple Cancer Gene Pathways 201

p53 Activation is Triggered by Damaged or Incompletely Replicated Chromosomes 204
p53 Induces the Transcription of Genes that Suppress Cancer Phenotypes 209
The MDM2-p53 Feedback Loop 211
The DNA Damage Signaling Network Activates Interconnected Repair Pathways 213
Inactivation of the Pathways to Apoptosis in Cancer 214
RB and the Regulation of the Cell Cycle 218
Several Cancer Gene Pathways Converge on Cell Cycle Regulators 220
Many Cancer Cells are Cell Cycle Checkpoint-Deficient 223
Overview: Dysregulation of Cancer Gene Pathways Confers Selective Advantages 224

6 Genetic Alternations in Common Cancers 227

Cancer Genes Cause Diverse Diseases 227
Cancer Incidence and Prevalence. 228
Lung Cancer. 229
Prostate Cancer 231
Breast Cancer 233
Endometrial Cancer 235
Lymphoma 237
Bladder Cancer 238
Melanoma of the Skin 240
Ovarian Cancer 242
Cancer of the Kidney 243
Leukemia 244
Pancreatic Cancer. 245
Cancers of the Oral Cavity and Pharynx 247
Cancer of the Uterine Cervix 248
Thyroid Cancer 250
Stomach Cancer 252
Brain Tumors 253
Liver Cancer 255

7 Cancer Genetics in the Clinic 259

The Uses of Genetic Information 259
Elements of Cancer Risk: Carcinogens and Genes 260
Identifying Carriers of Germline Cancer Genes 260
Altered Genes as Biomarkers of Cancer 262
Detecting Early Cancers via Gene-Based Assays 265
The Majority of Current Anticancer Therapies Inhibit Cell Growth 268
Molecularly Targeted Therapy: *BCR-ABL* and Imatinib 269
Clonal Evolution of Therapeutic Resistance 271

Allele-specific Cancer Therapy: Gefitinib 273
Antibody-Mediated Inhibition of Receptor Tyrosine Kinases 275
Targeting Death Receptors: TRAIL 276
Customized Cancer Therapy 277

Appendix **281**

Index **319**

Chapter 1
The Genetic Basis of Cancer

The Cancer Gene Theory

The human body is composed of a multitude of different cell types and tissues. Cancers can arise from all of these. What we broadly call cancer is actually a diverse spectrum of human diseases, a few of which constitute a mere nuisance while most others are deadly. The most common cancers in adults are *carcinomas*, derived from epithelial cells that line body cavities and glands. *Sarcomas* arise from mesenchymal tissues. Melanomas, retinoblastomas, neuroblastomas and glioblastomas are derived from dividing cells in the ocular retina, neurons and neural glia respectively. Lymphomas and leukemias, sometimes referred to as the liquid tumors, arise in the tissues that give rise to lymphoid and blood cells. All of these diseases will be collectively referred to as 'cancer' throughout this book. The rationale for this simplification is that all of these diverse diseases have a single root cause.

Cancer is caused by altered genes. The simplicity of this statement might be surprising, given the complexity of cancers. These diseases have many contributory factors and innumerable clinical manifestations. Nonetheless, there is an elemental concept that underlies this complexity. The tools of genetics have been used to systematically examine how cancers arise. Cancer researchers have pinpointed specific genes that are altered and demonstrated how these genetic changes cause tumors to grow in normal tissues. From decades of productive study, a theory has emerged that is both unifying and useful. Throughout this text, the assembled principles of cancer genetics will be referred to as *the cancer gene theory*. The cancer gene theory has provided a framework for understanding how both hereditary and environmental factors contribute to cancers. As will be described in the chapters that follow, this powerful theory will form the basis for new strategies for cancer prevention, detection and treatment.

The discovery that cancer is a genetic disease stands as one of the great triumphs of modern biomedical science. To put the importance of the cancer gene theory and its potential impact on public health in perspective, it may be useful to consider another epochal theory that preceded it: *the germ theory*. There are several notable similarities between infectious diseases and cancers in terms of how they were perceived by physicians of the early nineteenth century. Both types of diseases

F. Bunz, *Principles of Cancer Genetics*.

© Springer 2008

were common and fearsome ailments, shrouded in mystery and superstition. The underlying mechanism for each was essentially a black box. Both kinds of diseases were attributed to many different causes and were generally intractable to available forms of treatment.

The studies of Louis Pasteur and his contemporaries in the mid nineteenth century informed the germ theory, and thereby caused a revolutionary change in the way that infectious diseases were perceived. The idea that germs are the root cause of the broad range of what we now call infectious diseases created a scientific paradigm that eventually ushered in an age in which the causes, and eventually cures, of distinct infectious diseases could be systematically discovered and developed.

Infectious disease remains a complex entity, but the germ theory provides a simple framework for understanding how these diseases arise and how they might be catagorized and treated. A broadly diverse group of germs infect the various tissues of the body and respond to different classes of therapeutic compounds. Individuals vary in their susceptibility to different germs. Nonetheless, the germ theory that explains the underlying disease process provides a clear path to the understanding of any infectious disease.

The revolution in infectious disease research foreshadowed a similar breakthrough in cancer research that occurred a century later. The discovery of the molecular essence of the gene by James Watson, Francis Crick and their collaborators and the subsequent cracking of the genetic code opened the door to the explosion in molecular biological research in the latter part of the twentieth century. This enormous and productive effort has yielded the precise identification of genetic alterations that directly drive tumorigenesis, the process by which cancers arise, progressively grow and spread. Preventive and therapeutic anticancer measures based upon the cancer gene theory are at the early stages of development and hold great promise for the future.

The pioneers behind the germ theory showed that despite the complexity and diversity of infectious diseases, the underlying etiology of these diseases was relatively simple in concept. Simple concepts can be extremely powerful. Indeed, the germ theory forms the foundation for all modern attempts to classify, diagnose and treat the myriad diseases that are caused by infectious agents. This research continues today. A direct analogy between infectious disease and cancer is bound to be imperfect, and yet the similarity of the essential concepts is illustrative. As germs cause infections, cancer genes are the agents that drive cells to form tumors.

Cancers are Invasive Tumors

A *neoplasm* (literally ‘a new growth’) is any abnormal new growth of cells, whereas a *tumor* is a neoplasm that is associated with a disease state. Tumors are diseases in which a population of genetically related cells has acquired the ability to proliferate abnormally. The term ‘cancer’ simply defines those tumors

which have acquired the ability to invade surrounding tissues composed of normal cells. The distinction between a *benign* and a *malignant* tumor is solely based on this invasive capacity. If an invading malignant tumor reaches a blood or lymphatic vessel, a cancer can *metastasize* and grow in distant tissues. The ability of malignant cancers to disrupt other tissues and thereby spread is what makes them lethal.

As will be described in the following sections, tumors are thought to initially arise from a single, genetically altered cell. The growth of a tumor from the progeny of this one cell is a process known as *tumorigenesis*. As tumors grow from small, benign lesions to malignant and then metastatic cancers, the cells that compose these tumors change genetically and thereby acquire new properties. The acquisition of cancer genes underlies the process of tumorigenesis.

Cancer is a Unique Type of Genetic Disease

The best known, or classical genetic diseases are typically *monogenic* in nature, that is, they are caused by a single faulty gene. Some genetic diseases are relatively straightforward, in that their incidence is easily predicted by the Mendelian laws of inheritance. In such cases, inheritance of a gene defect is both necessary and sufficient to cause disease.

Sickle cell anemia is an example of a classical genetic disease. Disease is directly caused by a single alteration in the gene, *HBB*, that encodes beta globin, a subunit of hemoglobin. The protein encoded by this disease gene is relatively insoluble and can come out of solution under conditions of low oxygen tension, causing red blood cells to adopt the shape of a sickle and become nonfunctional. Anemia and vascular blockage are caused by the altered properties of the sickled red blood cells. There is an environmental component to acute illness, in the sense that a period of local oxygen deprivation is required to initiate the pathological process, but the underlying cause is clearly the disease gene. The pattern of inheritance of sickle cell anemia, like that of all monogenic diseases with high penetrance, is simple and can be predicted by the rules described by Mendel.

Like sickle cell anemia, cancer can be inherited as a monogenic trait. Large, extended families have been identified in which individuals in multiple generations develop related types of cancer at a high rate. Such families have been used to define cancer syndromes and to isolate the genes that underlie cancer susceptibility. While inherited cancer syndromes have provided a wealth of genetic information they are also relatively rare.

The majority of cancers that affect the human population cannot be predicted by the simple principles of Mendelian inheritance. The genes that cause cancer are not most commonly inherited, but rather are spontaneously acquired. Cancer is unique among genetic diseases in this regard. While genes that cause the classic genetic diseases are passed from generation to generation in a predictable way, cancer genes can be acquired in a number of additional ways.

What are Cancer Genes and How are They Acquired?

A cancer gene can be defined as a variant of a gene that increases cancer risk, or promotes the development of cancer. Cancer genes are distinct *alleles* of normal genes that arise as a result of mutation.

From the genetic perspective, there are two types of cells in the human body. *Germ cells* are the cells of the reproductive system that produce sperm in males and oocytes in females. *Somatic cells*, derived from the Greek work for body, *soma*, are all other cells exclusive of the germ cells. Cancer genes that arise in the germ cells are said to be in the *germline*. Individuals who inherit germline cancer genes will carry a germline cancer gene in every cell, somatic cells and germ cells alike. Such individuals are aptly known as *carriers*. In contrast, cancer genes that arise in somatic cells are not passed on to subsequent generations.

Tumors progressively acquire cancer genes as they grow. The mutations that define cancer genes can be acquired in three ways: (1) inheritance via the germline, (2) spontaneously via somatic mutation, and (3) via viral infection.

Inherited, germline cancer genes cause a small but significant fraction of human cancers. Depending on the cancer type, between 0.1% and 10% of cancers can be directly attributed to heredity. Several important cancer genes that are present in the germline of cancer prone families cause well-known cancer syndromes. The inheritance of such alleles greatly increases the probability that an individual will develop cancer. The likelihood that an allele carrier will develop cancer defines the *penetrance* of that allele. In some cases the penetrance of an inherited allele is so high that preemptive surgical treatment is indicated. Other germline cancer genes, many of which remain undiscovered, are likely to make smaller contributions to overall cancer risk. The means by which inherited genes cause cancer predisposition will be described in Chapters 3 and 4.

In the majority of cancers, the cancer genes that underlie tumorigenesis arise spontaneously by *somatic mutation*. Somatic mutation is a term that describes both a process, the spontaneous acquisition of a mutation in a non-germ cell, and a product, which is that genetic alteration. Somatic cells that spontaneously acquire cancer gene mutations are the precursors of cancers.

Both germline mutations and somatic mutations alter a normal gene and cause a new, mutant allele of that gene. Not all mutant genes acquired via the germline contribute to cancer risk, nor do all somatic mutations cause cancer. Indeed, the majority of genes and gene mutations do not appear to be associated with cancer.

The third way that an individual can acquire a cancer gene is by viral infection. This is a much less frequent mode of cancer gene acquisition and appears to be restricted to a relatively limited number of cancer types. As will be illustrated in sections that follow, viruses do play an important role in a significant number of common cancers. In most of these cancer types the contributory viruses do not actually carry or transmit cancer genes, but alter the environment in which cancer genes are propagated.

Mutations Alter the Human Genome

Somatic mutations are not heritable, while germline mutations arising in the germ cells that produce sperm and oocytes are passed vertically from generation to generation. Regardless of how they arise, there are a number of different types of DNA mutations that can alter the structure and function of a gene. When such a change occurs, a new variant, or allele of that gene is created. Small mutations that affect a relatively short region of DNA are typically detected by DNA sequencing, while larger mutations can be visualized by microscopy (see Fig. 1.1).

Mutations are typically categorized by the type and extent to which the DNA sequence is changed. *Single base pair substitutions*, often referred to as *point mutations*, simply change one base pair (bp) to another. More extensive mutations cause loss of DNA sequences or insertions of new DNA sequences. Deletions and insertions of 20 bp and less are typically called *microdeletions* or *micro-insertions*, respectively, while larger losses are termed *gross deletions* or *gross insertions*. This latter type of alteration can span many thousands of bp. Still larger scale processes can result in chromosome breaks that give rise to chromosomal translocations (see Fig. 1.2), deletions and inversions. Changes in chromosome structure within the microscopic size range are known as cytogenetic abnormalities. These large chromosomal rearrangements are rarely if ever transmitted via the germline in humans, and typically arise somatically.

The spectrum of known mutations is diverse. Largely uncharacterized mutational processes can result in DNA sequence inversion or in complex regions that

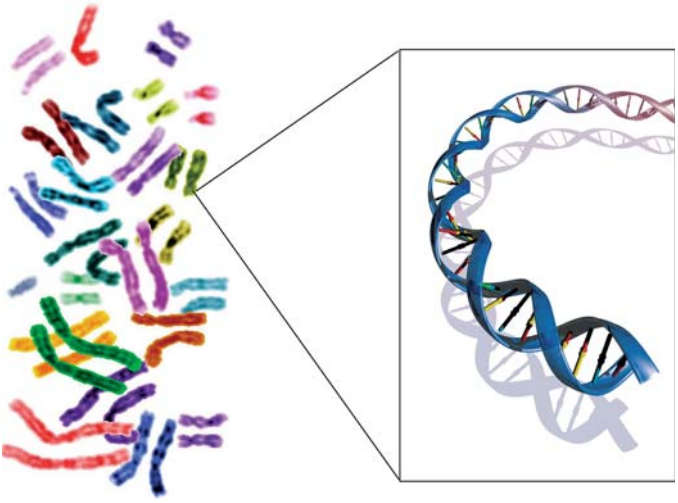

Fig. 1.1 Genetic alterations, great and small. The genetic alterations that underlie cancer can affect whole chromosomes (left), and therefore be detectable by cytogenetic methods. Small genetic alterations that affect individual DNA bases (right) are detected by molecular methods, including DNA sequencing. (Courtesy of the National Human Genome Research Institute.)

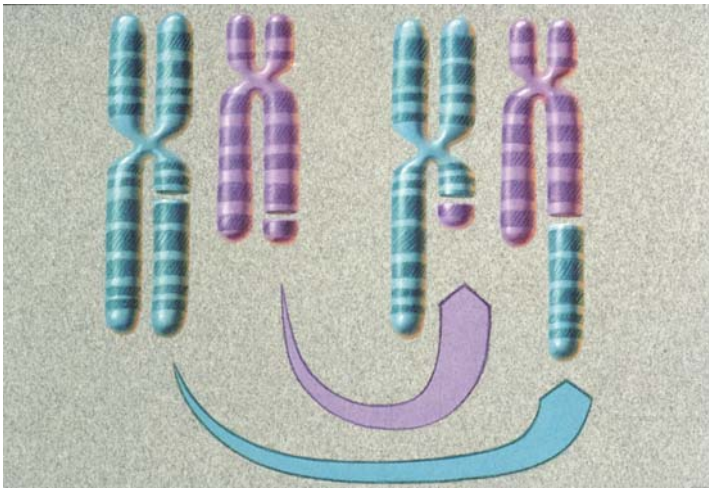

Fig. 1.2 Chromosomal translocation. The exchange of parts between nonhomologous chromosomes is known as a translocation. A balanced exchange between two chromosomes, as depicted in this example, is known as a *reciprocal translocation*

show evidence of both insertion and deletion. Short repetitive sequences can expand in tandem arrays. Long tracts of mononucleotide sequence (e.g. a tract of A residues that happens to have 57 A residues in a row, denoted A_{57}) can expand and contract. These different processes can alter genomic DNA sequences in virtually every imaginable way.

Understanding how mutations occur is critical to understanding the process of cancer. While much has been learned in this area, the origin of many mutations is incompletely understood. In some cases, there is considerable information as to the mechanisms by which some types of mutations occur. A significant fraction of all single base pair substitutions arise as a result of a normal cellular process called *DNA methylation*. Alterations occurring in mononucleotide tracts can often be attributed to defects in the processes by which genomic DNA is replicated and repaired. These specific mechanisms will be described more extensively in later sections.

In the case of gross changes that result in large deletions, insertions and chromosomal rearrangements, a possible mechanistic clue is the repetitive DNA sequences that flank many characterized deletion breakpoints. A substantial portion of the human genome is composed of repetitive elements. The most abundant of these is the *Alu* repeat, which were originally characterized with the use of the Alu restriction endonuclease. *Alu* repeats are highly similar regions that are about 300 bp in length. This core sequence is similar to bacterial sequences that stimulate recombination by promoting DNA strand exchange between sequences that have a high degree of similarity, or are homologous. Such evidence suggests that *Alu* repeats may represent hotspots for homologous recombination, which could theoretically create deletions and other types of large chromosomal rearrangements.

Regardless of their origin, many different alterations in genomic DNA can convert normal genes to cancer genes. Not all genomic alterations will contribute to cancer. In fact, only a small number of mutational events and rearrangements have been shown to promote tumorigenesis. The extent to which a mutation will give rise to cancer gene depends on the gene mutated, the region in which that gene is mutated and the precise nature of the mutation.

Genes and Mutations

How do mutations convert normal genes into cancer genes? To understand these critical events, it is useful to first review the basic elements of gene structure and function.

The gene defines a functional unit of heredity. The location of a gene is known as its locus. Genetic loci can encompass a chromosomal region that spans anywhere from fewer than 10^3 to greater than 10^6 bp, with a mean size of approximately 5×10^4 bp. The information content of a gene rests in the sequence of the four DNA bases: the purines adenine (**A**) and guanine (**G**) and the pyrimidines cytosine (**C**) and thymidine (**T**). Expression of a gene is determined by the rate at which the DNA at that locus is transcribed into RNA, and at which the RNA transcript is processed and translated into protein.

Genes that encode proteins are composed of several basic elements (see Fig. 1.3). The region of the gene that defines the protein that is ultimately expressed is the *open reading frame* (ORF). The ORF is a stretch of triplet DNA base pairs, or *codons*, that encode the distinct amino acids sequence of the expressed protein. ORFs are typically spread among multiple exons. During processing of the initially transcribed RNA, the exons are spliced together. Further processing usually includes the polyadenylation of the 3′ end of the transcript and creates a mature messenger RNA (mRNA).

Upon examination of the overall structure of a typical protein-coding gene, one would predict that alterations in different regions might have very different consequences. The ORF contains the information that is eventually translated into protein primary structure. Accordingly, mutations in these regions carry a relatively high probability of a functional consequence. In contrast, because introns are spliced out of RNA transcripts during processing, mutations within these regions would be predicted to have less potential impact.

Mutations within exons can alter gene function and ultimately affect protein structure (see Fig. 1.4). The most obvious way in which a mutation can directly alter gene function is by changing one or more codons, and thereby changing the amino acid sequence of the encoded protein. Mutations can also alter the way that RNA is spliced. Correct RNA splicing is dependent on the presence of short *splice donor, splice acceptor*, and *branch point* consensus sequences. Mutations within these splice sites can result in exon skipping or otherwise aberrantly spliced RNA species. Nearly 10% of all human gene mutations responsible for genetic diseases cause aberrant RNA splicing.

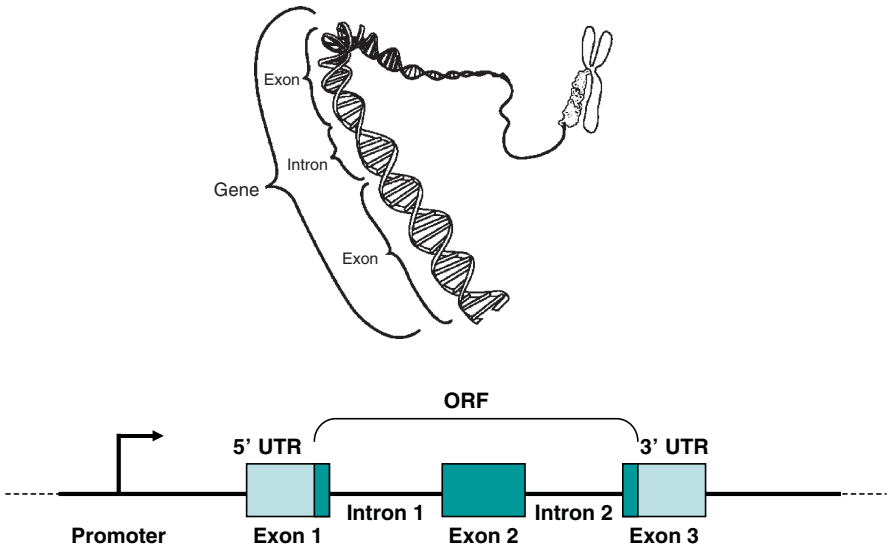

Fig. 1.3 Elements of a gene. A gene is composed of expressed regions known as exons and intervening sequences known as introns. In the chromosome, these elements are extensively packaged (top panel). The best understood genes encode proteins. As shown in a linear representation (lower panel), exons contain a protein-coding region known as an open reading frame (ORF). Flanking the ORF are the 5′ and 3′ untranslated sequences (UTRs). Expression of the gene requires the activity of a cis-acting promoter

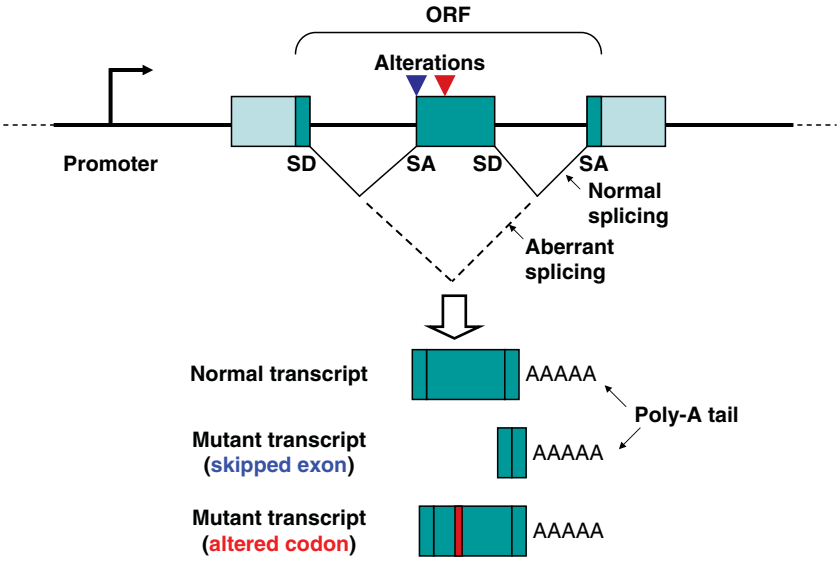

Fig. 1.4 Mutations can alter gene transcripts. The majority of the mutations that create cancer genes alter codons (example shown in red). About 10% of mutations interfere with RNA processing by disrupting the splice donor (SD) or splice acceptor (SA) consensus sequences. Such mutations (example shown in blue) cause aberrant RNA splicing that can result in exon skipping

While mutations that occur within introns generally appear to be of no discernable functional consequence, some unusual mutations in introns have been shown to affect gene function. In rare instances, mutations within introns activate *cryptic splice sites*, essentially generating new splice sites that then lead to the production of aberrant RNA species. Other intron mutations have been shown to alter splicing efficiency in ways that are not well understood.

Mutations in promoter elements, transcriptional initiation sites, intiation codons, polyadenylation sites and termination codons have also been shown to alter gene function. All of these together account for less than 2% of all mutations known to cause human disease, including cancer.

The gene concept has expanded in recent years, as new technologies have been used to globally monitor transcription of RNA. The notion that discrete genes produce distinct transcripts, which in turn uniformly give rise to biochemically active proteins has gradually given way to a broader view of what a gene actually is and does. Large-scale analysis of the genome and the RNA transcribed from it has revealed that there is a great deal more transcription in the cell than the one-gene one-protein model would predict. A transcriptional survey of the mouse has revealed that while 1–2% of the genome is spanned by groups of classical exons, an astounding 63% of the mouse genome is actively transcribed! The human genome is also pervasively transcribed. What is the function of all this RNA? The sheer mass of RNA produced, the energy required to produce it and the size and complexity of the genome suggest that there is an important functional component to regions of the genome outside what we would recognize as classical genes.

Are genomic regions that express non-coding RNA actually genes, that is, units of heredity? And do these genetic entities contribute to cancer? At this point, it is difficult to predict to what extent somatic and germline mutations in non-coding regions of the genome might contribute to cancer. It is important to note that in the many cases in which scientists have successfully found the underlying mutation that causes a genetic disease, virtually all of these mutations have affected proteins. Though the broadening of the gene concept has been exciting, the balance of evidence still suggests that disease-causing mutations predominantly affect protein coding regions of the genome. The role of non-coding regions in genetic diseases like cancer remains to be determined.

Genetic Variation and Cancer Genes

Humans are a genetically diverse population. The broad spectrum of phenotypic traits present within our species results from the genetic variation between individuals, much of which remains to be quantified. Indeed, these genetic differences underlie many of the characteristics that define us as individuals. Our unique set of genes contributes much to who we are and what we look like, and similarly contributes to our predisposition to disease.

By any measure, the human genome contains an enormous amount of information. Every nucleated cell of the human body contains about 3.5 picograms of DNA. The haploid maternal and paternal genomes within each diploid cell are composed of 3.4 billion base-paired nucleotides. As we will see in the chapters that follow, alterations in even a single base pair in a critical region of a regulatory gene can confer a risk of cancer so high that disease is essentially inevitable. Genetic alterations that appear miniscule can have significant consequences.

At the DNA sequence level, humans are more than 99% identical. The genetic differences between any two unrelated individuals lie in the small proportion of the human genome that is variable. Specific variations in the genome are known as *polymorphisms*. These can be in the form of single base variants, insertions, deletions, variations in repeat elements, and more complex rearrangements. The most common form of genetic variation between humans is the *single nucleotide polymorphism*, or SNP (pronounced 'snip'). When the genomic DNA sequences of two individual, homologous chromosomes are compared, SNPs occur, on average, every 1,000–2,000 bp. It is estimated that there are about 10–12 million different SNPs in the human population. Roughly 4% of SNPs occur in exons; most exons are within 5,000 bp of the nearest SNP. The multitude of possible SNP combinations that can occur in a given individual accounts for a large proportion of human genetic variation.

The genetic changes, known as mutations, that cause cancer thus must be evaluated against a background of significant diversity. How are cancer genes to be distinguished from more benign genetic variants? The answer to this question is rarely straightforward. The branch of genetics concerned with the statistical association of specific genetic variants with diseases has come to be known as molecular epidemiology.

Somatic mutations can readily be differentiated from SNPs by comparing cancer cells to normal cells from the same individual. In this clinic, this can be accomplished by the examination of biopsy samples that contain both cancer cells and the cells that make up the normal surrounding tissues. Unlike polymorphisms, somatic mutations arise spontaneously and will therefore only be present in the cancer cells. The extent to which a given somatic mutation actually contributed to development of a cancer is an unrelated but obviously important question that will be discussed in later sections.

More problematic is the evaluation of genetic variants that are present in the germline of individuals with cancer. All variants originally arose by mutation of the genome, but not all mutations cause cancer. Which variants contribute to disease and which are incidental? How can a cancer gene be distinguished from a benign variant? Most genetic variants, be they SNPs or other types of polymorphism, are largely unrelated to a person's risk of developing cancer. As will be described in the subsequent chapters of this book, the isolation of genetic variants that cause increased cancer risk has been a challenging undertaking marked by remarkable triumphs.

There are several clues that might indicate that a given SNP or other variation measurably affects the risk of cancer. One important parameter is the *allele frequency*.

It appears that inheritance plays a significant role in only a subset of all cancers. Most known cancer genes are acquired by somatic mutation rather than inheritance. These facts suggest that germline cancer genes should be relatively uncommon. Common SNPs probably do not impart large cancer risks. For example, if a SNP present in an individual from a cancer-prone family is also present in a large proportion of individuals that are not particularly predisposed to developing cancer, then that SNP is unlikely to define an important cancer gene.

The pattern of inheritance within a family pedigree is a critical criterion for identifying a cancer gene. A germline cancer gene would be expected to *cosegregate* with cancer predisposition. The allele suspected to be a cancer gene should be present in family members who develop inherited cancers, and absent in those that do not.

The location of a variation and the consequences of that variation on protein function are additional factors to consider. Mutations can occur anywhere in the genome. Many of these changes will have little obvious effect on gene function. In contrast, most known mutations that increase cancer risk have measurable effects on gene function or expression. Unlike the majority of mutations that occur in non-expressed regions of the genome, those that are known to contribute to cancer risk most often are located in or near exons and affect the structure and function of encoded proteins.

Much remains to be learned about how genetic variation contributes to cancer risk. Most of the inherited cancer genes discovered to date have a high penetrance and impart a significant predisposition to the development of cancer. Fewer low penetrance cancer genes are known, largely because such genes are more difficult to identify. Genes that may modify cancer risk in subtle ways are far more difficult to detect but may collectively cause a significant number of cancers. Extensive statistical analysis of compiled genetic information will be necessary to understand to a fuller extent to which cancer is inherited and facilitate the discovery of new cancer genes.

One approach to minimizing the confounding effects of population diversity is the extensive evaluation of small, well-defined human subpopulations. As will be described in Chapter 3, a number of important and highly informative cancer genes have been isolated, in part, because of their inheritance within defined ethnic groups.

The ideal population for epidemiological study is one in which disease is well documented over many generations and overall genetic variability is limited. The people of Iceland have been proposed as one such genetic resource. Iceland, a wealthy nation with universal access to healthcare, contains a population of about 300,000 individuals who can directly trace their ancestry to a relatively small number of founding individuals. The potential value of this enormous pedigree is underscored by the rights to this information that have been secured by a commercial entity. It remains to be seen whether the incidence of inherited cancers is sufficiently high, against a background variation that is sufficiently low, to provide the statistical power to identify novel cancer genes within the Icelandic population.

Which Mutations are Important in Cancer?

Not all mutations are equivalent. A mutation in a coding sequence is much more likely to result in a change in gene function than a change in an intron or a non-coding exon. Among the mutations that occur within coding exons, some have much larger effects than others. Some mutations result in no phenotypic effect while other changes can profoundly affect gene function and alter disease risk.

Some single base pair substitutions do not result any change to the encoded protein. The reason lies in the inherent degeneracy of the genetic code; many amino acids have several codons that are synonymous. Leucine, for example, can be encoded by six DNA triplets: CTT, CTC, CTA, CTG, TTA and TTG. A C→T change that results in a mutation of CTC to CTT will have no net effect. In this case, one leucine codon is simply converted to another. Such mutations are known as *silent mutations*, and are the most benign type of mutation in terms of disease risk. Mutations in the third codon position, also known as the *wobble* position, are least likely to result in an amino acid change.

A single base pair substitution that causes a codon change is known as a *missense* mutation. A C→A mutation would change CTT, the codon for leucine into ATT, which encodes isoleucine (see Table 1.1). In this case, a single base change results in a single amino acid change. A single base pair substitution can also change a codon that represents an amino acid into one of the termination, or STOP, codons, encoded in the DNA sequence by TAG, TAA, and TGA. Terminating mutations, also known as *nonsense* mutations, result in truncation of the open reading frame.

Table 1.1 The standard genetic code. The DNA codons are grouped with their corresponding amino acids (the single-letter amino acid designations are in parentheses). The degeneracy of the genetic code reduces the impact of many single nucleotide substitutions.

Second codon position		T	C	A	G
First codon position	**T**	TTT Phe (F)	TCT Ser (S)	TAT Tyr (Y)	TGT Cys (C)
		TTC	TCC	TAC	TGC
		TTA Leu (L)	TCA	TAA **STOP**	TGA **STOP**
		TTG	TCG	TAG **STOP**	TGG Trp (W)
	C	CTT Leu (L)	CCT Pro (P)	CAT His (H)	CGT Arg (R)
		CTC	CCC	CAC	CGC
		CTA	CCA	CAA Gln (Q)	CGA
		CTG	CCG	CAG	CGG
	A	ATT Ile (I)	ACT Thr (T)	AAT Asn (N)	AGT Ser (S)
		ATC	ACC	AAC	AGC
		ATA	ACA	AAA Lys (K)	AGA Arg (R)
		ATG Met (M)	ACG	AAG	AGG
	G	GTT Val (V)	GCT Ala (A)	GAT Asp (D)	GGT Gly (G)
		GTC	GCC	GAC	GGC
		GTA	GCA	GAA Glu (E)	GGA
		GTG	GCG	GAG	GGG

Missense mutations can have a large range of phenotypic effects. The effect of a missense mutation depends on both the relatedness of the original and mutated amino acids and the position of the change within the structure of the encoded protein. In our previous example, leucine and isoleucine are structurally very similar and have the same molecular weight. In many proteins, the substitution of a leucine for an isoleucine would have little demonstrable effect. In contrast, the mutation of GAG, which encodes glutamic acid, to a GTG codon for valine results is a change from a highly acidic to a hydrophobic amino acid. A single base change thereby causes the amino acid substitution that is the basis for the gross structural and functional changes in β-globin that underlie sickle cell anemia.

The position of an amino acid substitution within an encoded protein is also a key determinant of the extent to which a mutation can alter gene function. Protein structure is progressively defined by amino acid sequence (*primary structure*), by interactions between neighboring amino acids (*secondary structure*), by three-dimensional interactions between more distant peptide motifs (*tertiary structure*) and finally, by interactions between subunits of multiprotein complexes (*quaternary structure*). By definition, all missense mutations alter the primary structure. Some, but not all, missense mutations can also change the tertiary structure. Mutations that change amino acids that directly contribute to disulfide bonds, hydrophobic interactions and hydrogen bonds affect both secondary and tertiary protein structure and often result in dramatic functional changes. For proteins that function as catalytic enzymes, mutations near the substrate or cofactor binding domains can profoundly influence activity. Structural proteins, in contrast, are typically sensitive to mutation in regions involved in the critical protein–protein interactions that define their quaternary structure. In general, amino acid residues that are present in similar positions in homologous proteins from other species, and are therefore *evolutionarily conserved*, are more likely to have a functional impact when mutated.

Because an open reading frame is defined by a continuous array of triplet codons, any alteration to this invariant pattern will have significant effects. Thus, even small deletions and insertions can completely disrupt an open reading frame. If a deletion or insertion within an open reading frame involves any number of bp not divisible by 3, that alteration will result in a shift in the reading frame. *Frameshift mutations* invariably result in a new set of codons that encode an entirely unrelated series of amino acids in the 3′ direction (downstream) from the location of the mutation. Because the human genome is rich in the A:T bp that are present in stop codons, probability dictates that any given alternate reading frame resulting from a frameshift will have a termination codon within a short distance (see Fig. 1.5). Small insertions and deletions therefore typically result in a new coding sequence that encodes both random amino acids and a truncated protein product. The closer a mutation occurs to the 5′ end of an open reading frame that encodes the amino terminus of a predicted protein, the greater the effects on protein function.

Mutations that affect correct splicing of exons can often lead to aberrations such as *exon skipping* (see Fig. 1.4) and activation of cryptic splice sites. Such alterations will usually lead to a shift in the reading frame, with the same consequences as

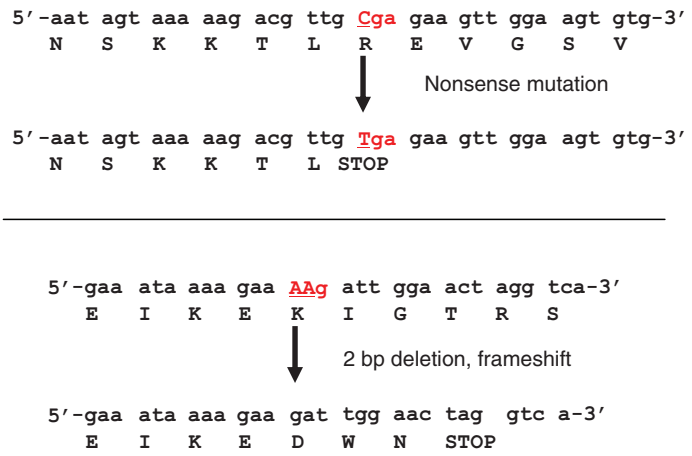
5′-aat agt aaa aag acg ttg Cga gaa gtt gga agt gtg-3′
N S K K T L R E V G S V
Nonsense mutation
5′-aat agt aaa aag acg ttg Tga gaa gtt gga agt gtg-3′
N S K K T L STOP
5′-gaa ata aaa gaa AAg att gga act agg tca-3′
E I K E K I G T R S
2 bp deletion, frameshift
5′-gaa ata aaa gaa gat tgg aac tag gtc a-3′
E I K E D W N STOP

Fig. 1.5 Truncating mutations. Nonsense mutations generate STOP triplets (upper panel). In this example, a C→T mutation (indicated in red) introduces a premature STOP. Insertions or deletions create frameshifts that contain premature STOP triplets (lower panel). In this example, the deletion of AA (indicated in red) results in a frameshift and the appearance of a premature STOP several codons downstream

other types of frameshift mutations. In the case in which the skipped exon contains a multiple of 3 bp, the spliced mRNA product will maintain the original reading frame, with the only consequence of the mutation being the loss of the amino acid positions encoded by the skipped exon.

Premature stop codons caused by nonsense or truncating mutations do not typically result in the expression of truncated protein because mRNA transcripts that contain nonsense codons are systematically and rapidly degraded. The multistep pathway that performs this surveillance function is known as *nonsense-mediated mRNA decay*. This process can distinguish between normal and premature stop codons. Nonsense-mediated mRNA decay is an evolutionarily conserved process that is thought to be a mechanism to eliminate mRNAs that encode for potentially deleterious protein fragments. It has been estimated that up to one quarter of all cancer mutations are of the type that could trigger nonsense-mediated decay, though the actual contribution of this pathway to the reduction of cancer gene expression remains to be determined.

In summary, nonsense mutations and truncating insertions and deletions have multiple consequences, including open reading frame alteration and truncation and suppression of expression by nonsense-mediated mRNA decay. It is straightforward to imagine how these effects, in combination, might totally result in the total inactivation of a gene. An allele that expresses no gene product, or encodes a gene with no activity, is known as a *null allele*. Less common genetic alterations can also cause null alleles. For example, a gross deletion sufficiently large to eliminate an entire open reading frame would create a null allele.

While many cancer-causing mutations cause the generation of null alleles, many seemingly minor genetic changes change normal genes into cancer genes. In fact,

the most common cancer-causing mutations involve small changes to the DNA sequence. As will become apparent, small genetic changes can carry large biological consequences.

Single Nucleotide Substitutions

The most common type of DNA mutation is the substitution of a single nucleotide. A mutation that substitutes a single DNA base for another is often referred to as a *point mutation*. (Although both base pairs are affected by a single nucleotide substitution, the base that is on the coding DNA strand is the alteration most commonly noted.)

A transition is a base change from one purine to another, or from one pyrimidine to another (e.g. C→T or G→A). A transversion is a change from a purine to a pyrimidine or vice versa (e.g. A→T or C→G). Given that there are four bases, a total of 12 different types of base substitutions are possible (see Fig. 1.6).

While each base can be mutated and replaced by any other base, some substitutions are much more common than others. The most frequent substitutions are C→T and G→A, which together account for nearly 50% of all single base substitutions. These rates are obviously much higher than would be expected by random chance. The reason for the unexpected overrepresentation of C→T and G→A base changes is the inherent mutability of the CG dinucleotide (usual written as CpG to emphasize the 5′ → 3′ orientation of C to G).

CpG dinucleotide sequences are frequently the target of a chemical modification known as *DNA methylation*. The covalent modification of the cytosine ring by a family of enzymes called DNA methyltransferases converts cytosines that are located 5′ to guanosines to 5-methylcytosine (5mC). 5mC has a propensity to undergo *deamination* to become uracil, which in turn becomes a thymidine during the next round of DNA replication if the deaminated base has not been repaired (see Fig. 1.7). The resulting C→T transition is mirrored by a corresponding G→A transition on the complementary DNA strand. As a result of methylation and subsequent deamination, CpG dinuceotide sequences have been progressively lost from

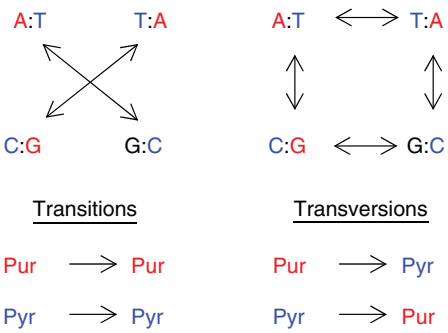

Fig. 1.6 Transitions and transversions. A total of 12 distinct base changes are possible

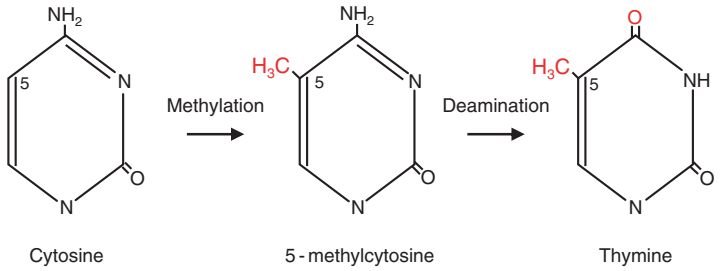

Fig. 1.7 Endogenous methylation causes a C→T transition. DNA methyltransferases convert C to 5-methylcytosine (5mC). This reaction occurs preferentially at CG dinucleotides. The ring containing 5mC is converted to a T by loss of the NH_2– group, a chemical reaction known as *deamination*

the human genome over the course of many generations. Thus, the hypermutability of CpG sequences has led to a relative paucity of CpG sites in the human genome.

The stochastic transitions caused by CpG mutation are a source of significant variation in the human genome. CpG mutations in germ cells that give rise to sperm and oocytes can result in germline mutations. Somatic mutations can also occur via this process. Because of the inherent mutability of this dinucleotide, regions of the genome that are CpG rich are often called mutation *hotspots*.

While the inherent hypermutability of CpG dinucleotides causes mutations that can convert normal genes to cancer genes, other processes can also cause single nucleotide substitutions. Mutations can arise from the process of DNA replication itself via base misincorporation by the replicative DNA polymerase complexes. There are several mechanisms by which the DNA replication apparatus is thought to cause mutations:

Slipped Mispairing in Mononucleotide Tracts. Runs of identical bases can adversely impact DNA replication fidelity. At the replication fork, discontinuous synthesis of the lagging strand is mediated by the iterative extension of primers. One mechanism of mutagenesis is thought to arise from transient misalignment of the primer-template that results from the transient looping out of a base on the template strand (see Fig. 1.8). A base is thus misincorporated into the primer strand, resulting in a mismatch. If the mismatch is repaired in favor of the strand with the misincorporation, a mutation results. Known as the Slipped Mispairing Model, devised by Thomas Kunkel, this mechanistic explanation for replication-associated mutagenesis is supported by an observed bias in the identity of the mutated base to a flanking base within open reading frames. For unknown reasons, this bias is limited to the first two codon positions. In principle, slipped mispairing could also generate a one base insertion or deletion, depending on the primer-template misalignment and repair of the mismatch. It is unclear to what extent this actually happens.

Deoxynucleotide availability. DNA synthesis depends on the availability of raw materials, the four deoxyribonucleotides (dATP, dCTP, dGTP and TTP, collectively

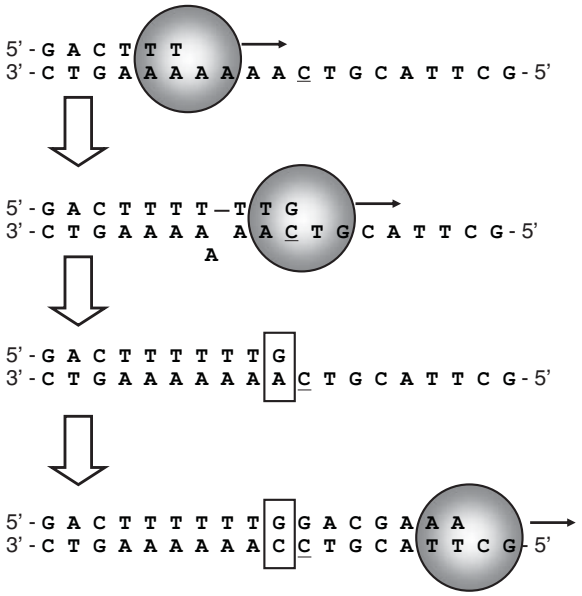
5'-G A C T T T
3'-C T G A A A A A A A C T G C A T T C G-5'
5'-G A C T T T T—T T G
3'-C T G A A A A A A C T G C A T T C G-5'
A
5'-G A C T T T T T T G
3'-C T G A A A A A A A C T G C A T T C G-5'
5'-G A C T T T T T T G G A C G A A A
3'-C T G A A A A A A C C T G C A T T C G-5'

Fig. 1.8 Slipped mispairing in an A_7 tract. In this example, a DNA polymerase holoenzyme complex (shown as a sphere) encounters a tract of seven 'A' nucleotides. The looping-out of an 'A' on the template strand causes a transient misalignment of the primer and template DNAs. A 'G' is thus misincorporated into the primer strand at a position that would correctly be occupied by a 'T'. The realignment of the primer-template strand reveals a G:A mismatch. During DNA repair, the replacement of the 'A' on the template strand would represent a mutation

referred to as dNTPs). The mobilization of dNTPs during DNA replication or DNA repair is highly regulated and concentrations of dNTP pools tightly controlled. The fidelity with which DNA polymerases replicate a template DNA strand is highly sensitive to dNTP levels. The probability of misincorporation of a base will depend partly on the ratio of the correct dNTP to the three incorrect dNTPs available to the DNA polymerase. After a misincorporation has occurred, the efficiency with which it is excised before additional synthesis proceeds depends partly on the concentration of the next correct dNTP to be incorporated, which if high, will favor mismatch extension. Thus, alterations in dNTP proportions or total dNTP concentration can both affect DNA replication fidelity.

Stalled replication forks. The rate of base misincorporation can change dramatically if the progress of the replication fork is impeded. Short DNA sequences that have been identified as disproportionate targets of mutation are thought to directly cause the replication fork to stall or pause. For example, the sequences TGGA and TCGA are mutated at twice the rate that would be expected by chance alone, and this sequence also resembles a site at which DNA polymerase α has been shown to transiently arrest.

Low fidelity DNA repair. The DNA–polymerase complexes responsible for the repair of damaged DNA have a significantly lower fidelity, that is, are much more

error-prone, than replicative DNA polymerases. Switching between these polymerases during DNA repair processes results in an overall increase in misincorporation. This low fidelity, and correspondingly higher rate of base misincorporation, is thought to be a significant mechanism by which environmental agents can cause mutations.

Gene Silencing by Cytosine Methylation: Epigenetics

The CpG dinucleotides that are the targets of DNA methyltransferases are distributed asymmetrically throughout the genome. Most regions of the genome have been depleted of CpG sites by spontaneous deamination. However, discrete regions known as *CpG islands* retain the number of CpG dinucleotides that would be predicted to occur randomly. CpG islands, which range in size between 0.4 and 5 kb are often associated with gene promoters.

The methylation of CpG islands near gene promoters is associated with the downregulation of gene expression, a phenomenon also known as *gene silencing*. There is a striking difference in the methylation patterns in normal cells and cancer cells. Most gene promoters in normal cells are unmethylated and therefore capable of driving transcription. In contrast, many promoters in cancer cells are *hypermethylated*, with their corresponding genes thus transcriptionally silenced. Patterns of CpG DNA methylation, known as *epigenetic* alterations, can be inherited in a process known as *imprinting*.

CpG methylation is a cause of two types of heritable changes: genetic alterations (C→T transitions) and epigenetic alterations (gene silencing). Aberrant CpG methylation and gene silencing represent an alternative mechanism to genetic alteration. Many known cancer genes, defined by mutations, are among the genes found to be reversibly silenced via hypermethylation in cancer cells. As a result, epigenetic mechanisms have been proposed to account for many of the phenotypic abnormalities that arise during tumorigenesis, including dysregulated cell growth, cell death and genetic instability. The overall contribution of epigenetic alterations to human cancer remains to be definitively determined, but the aberrant CpG methylation patterns found in cancer cells are an intriguing observation.

Environmental Mutagens, Mutations and Cancer

It is well known that agents in the environment can cause cancer. Exposure to certain agents results in a clear and potent increase in risk for the development of common cancers. The respective contributions of tobacco smoke and sunlight to lung and skin cancers are excellent examples of this cause and effect relationship.

How do the incontrovertible relationships between cancer and the environment relate to the cancer gene theory? Part of the answer is that some environmental

agents are *mutagens*, that is, exposure to these agents increases the rate at which specific mutations appear. The cancer gene theory thus explains one way that environmental factors can contribute to cancer. Mutagens cause mutations that cause cancer.

For the purposes of illustration, consider a single gene, *P53*, and the environmental factors that can contribute to its mutation. *P53* is mutated in many cancers and is the most intensively studied cancer gene. As will be described in later chapters, insights related to *P53* have been a pillar of the cancer gene theory. In this section, the focus will be on the ways that environmental agents can cause the mutation of *P53* and thereby create a cancer gene. The biology of the *P53* gene and the ways that *P53* mutations cause cancer will be considered at length in later chapters. It is important to note that the mutagens discussed below alter other genes in addition to *P53* and that *P53* is mutated by additional processes that remain incompletely understood.

(The mutations indicated hereafter are described in reference to the base change that occurs on the coding, or sense, DNA strand. For example, a C→T transition on the sense strand is necessarily coincident with a complementary G→A transition on the antiparallel, antisense strand.)

Tobacco smoke. The relationship between cigarette smoking and cancer is one of the most clearly defined examples of the carcinogenic potential of environmental agents. Smokers have a tenfold greater risk of dying from lung cancers and this risk increases to 15- to 25-fold for heavy smokers. Only 5–10% of all lung cancers occur in patients that have no prior history of cigarette smoking. In addition to the well-known causative association between smoking and lung cancer, smoking is also a significant risk factor for a number of other cancers, including head and neck cancer and urinary bladder cancer.

Polycyclic aromatic hydrocarbons generated by the incomplete combustion of organic material during smoking are strongly implicated as the carcinogenic component of tobacco smoke. Among these, benzo[a]pyrene is by far the best studied. After ingestion, benzo[a]pyrene is metabolically altered to benzo[a]pyrene diol epoxide, or BPDE, by the P450 pathway. There are several isomers of this highly mutagenic metabolite that are formed during this process. The mucosal linings of the lungs, head and neck and the urinary bladder epithelia are all highly exposed to BPDE in smokers, further underscoring the relationship of these tissues to the cancer-causing effects of tobacco smoke.

BPDE binds directly to DNA and forms four structurally distinct covalent adducts at the N2 position of guanine (see Fig. 1.9). The N2-BPDE-dG adducts constitute a significant barrier to DNA replication forks. The repair process that deals with such lesions results in a high proportion of G→T transversion mutations. The factors that determine whether a given N2-BPDE-dG adduct will give rise to a single base pair substitution are complex, and partially depend on the stereochemistry of the specific adduct and the sequence and methylation status of neighboring bases.

That BPDE contributes to smoking-related cancer by causing mutations is supported by the types of *P53* mutations actually found in lung cancers. The *P53*

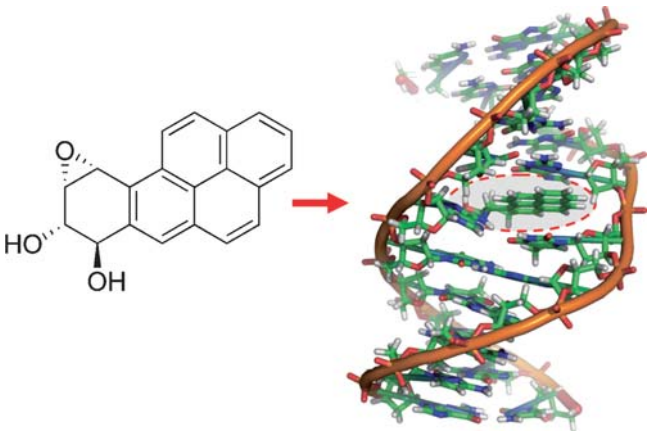

Fig. 1.9 BPDE forms a DNA adduct. The BPDE molecule (left) intercalates in the DNA double helix (right) and covalently bonds to a guanine residue at the N2 position. (Illustration by Richard Wheeler. Data from Pradhan et al. *Biochemistry* (2001) **40**, 5870–5881.)

mutations commonly found in lung cancers are not found at random, but rather at known hotspots, or regions within the *P53* coding sequence that are mutated at high frequency in large numbers of lung cancers that have been examined. The base positions within the *P53* open reading frame at which BPDE preferentially forms adducts overlaps significantly with known mutation hotspots, suggesting that BPDE directly causes the mutations that contribute to lung cancer.

Ultraviolet (UV) light. Sunlight is the main cause of basal and squamous cell cancers of the skin. The UV-B component of sunlight, encompassing wavelengths 290–320 nm in the electromagnetic spectrum, is a mutagen that causes two types of alterations to adjacent pyrimidines: *cyclobutane dimers* and *pyrimidine (6–4) pyramidone photoproducts* (see Figs. 1.10 and 1.11). Most pyrimidine photoproducts are repaired by a process known as nucleotide excision repair, which will be described in detail in Chapter 4. Failure of this repair mechanism results in a single nucleotide substitution.

Skin cancers that arise in sun exposed areas have frequent mutations in the *P53* gene and in other genes. Most of the mutations observed are C→T single base transitions with a significant number of CC→TT double base changes. The UV-B-induced photoproducts largely affect pyrimidines that are adjacent to other pyrimidines. In cases of the C→T single base transition, there is a significant bias towards mutation of C bases that occur in CpC dinucleotides. The CC→TT double base mutations observed occur most commonly in the context of the triplet sequence CCG. The CpG dinucleotide is frequently methylated in the genome, suggesting that the double base changes observed probably result from the unique resolution of a photoproduct next to a methyl-cytosine base. These base changes are unique to UV-B-mediated mutagenesis, and are often referred to as the *UV signature*.

Ionizing radiation (IR). Human tissues are constantly bombarded with high-energy subatomic particles. Sources of ionizing radiation in the environment are

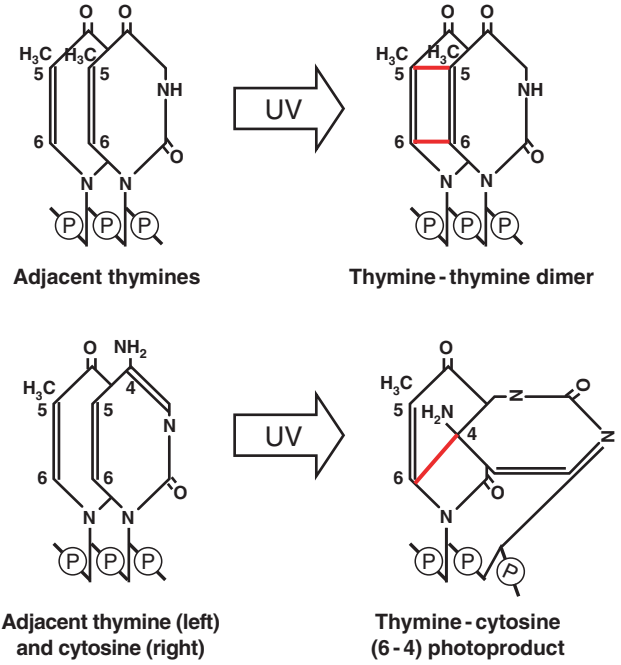

Fig. 1.10 Two predominant UV-induced DNA lesions. Formation of a cyclobutane thymidine dimer (top). Formation of a (6–4) photoproduct between an adjacent thymidine and cytosine (bottom). A significant degree of distortion of the phosphodiester DNA backbone is caused by (6–4) photoproduct formation

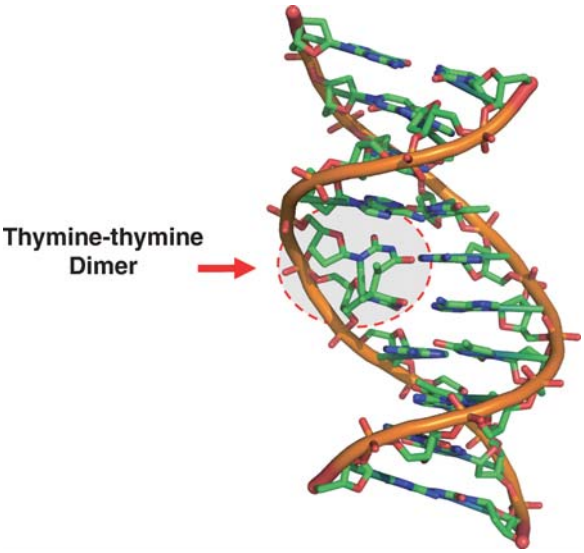

Fig. 1.11 Thymine-thymine dimer. This three-dimensional rendering of a thymine dimer reveals the local disruption of normal base pairing. (Illustration by Richard Wheeler. Data from Park et al. *Proc. Nat. Acad. Sci*. (2002) 99:15965–15970.)

both natural and anthropogenic. Depending on where they live and work, individuals encounter varying levels of radon gas that arises from the earth's crust and cosmic radiation that penetrates the atmosphere. Medical x-rays are a significant source of exposure for some people. Radioactive fallout from nuclear weapons and nuclear accidents are problematic in more restricted areas, most notably Hiroshima and Nagasaki in Japan and the region near Chernobyl in the Ukraine.

When a subatomic particle of sufficient energy passes through a cell, it leaves a narrow track of ionized molecules in its wake. A large proportion of these unstable molecules are reactive oxygen species. These unstable and highly reactive molecules disrupt the phosphodiester bonds that form the DNA backbone and often result in a double-strand DNA break. Agents that create double-strand DNA breaks are known as *clastogens*.

Ionizing radiation is a potent clastogen, but a significantly weaker mutagen. In other words, radiation causes many chromosomal breaks, but few of these resolve into a stable mutation that can be propagated by cell division. There are two known ways in which double-strand DNA breaks are repaired: *non-homologous end joining* (NHEJ), in which the two free ends of a broken chromosome are essentially fused back together, and *homologous recombination*, where the intact sister chromatid is used as a repair template. While homologous recombination uses extensive regions of sequence homology to align the damaged strand to the repair template, NHEJ exploits very short regions of incidental sequence similarity, termed *microhomologies*, to bring together and repair the damaged ends. Both of these processes can reconstruct the original sequence in the majority of cases. NHEJ is the more error-prone of the two repair mechanisms due to erroneous pairings that occur by chance. Slippage between regions of microhomology contributes to NHEJ errors, particularly in mononucleotide repeat tracts. End processing that occurs during NHEJ can also contribute to errors. Despite these sources of error, NHEJ has an error rate of only 1%. The predominant mutation caused by ionizing radiation is the microdeletion as would be expected if slippage during NHEJ was the principal mechanism involved.

Single nucleotide substitutions can be detected in radiation-associated cancers, though there is more limited information as to how these arise. Exposure to high doses of ionizing radiation has been shown to correlate with the appearance of several cancers, including cancer of the liver and basal cell cancer of the skin. Analysis of the *P53* gene in liver cancers associated with radiation exposure reveal a substantial number of single base alterations that affect the expressed protein. The largest proportion of these is the C→T transition, predominantly occurring at non-CpG sites. This negative bias against the CpG dinucleotide implies that the observed transition is less likely to result from the accelerated turnover of methylated cytosine, but rather results from the direct modification of bases by a direct effect of radiation. It is thought that direct oxidative modifications to cytosine might contribute to the later appearance of point mutations.

P53 mutations are also found in skin cancers from individuals exposed to high levels of radiation. Assessment of survivors of the Japanese atomic bomb blasts has provided valuable clues to the nature of radiation induced single nucleotide

substitutions. In these individuals, the etiology of the skin cancer can be inferred from the location of the lesion. Skin cancers that occur in areas unexposed to sunlight are presumed to be associated with ionizing radiation exposure. The UV-associated basal cell cancers from these individuals contained the UV signature mutations described in the preceding section. The lesions attributable to ionizing radiation, in contrast, had *P53* mutations that were C→T transitions at predominantly at non-CpG sites, similar to those observed in ionizing radiation-associated liver cancers.

Aflatoxin B1. Dietary exposure to aflatoxins are a significant risk factor for the development of liver cancer. Aflatoxins are produced by fungi commonly found in regions of southeast Asia and sub-Saharan Africa that grow on foods such as corn, rice and peanuts. Liver cancer is also endemic to these areas (see Chapter 6). A subtype of aflatoxin, known as aflatoxin B1 (AFB1), is a potent carcinogen that can induce liver cancer in animal models. Whereas the environmental agents previously discussed cause an array of different, though structurally related, DNA base changes, exposure to AFB1 has been found to result in a single, unique alteration to the *P53* gene. In more than 50% of tumors that arise in areas with high levels of environmental AFB1, a G→T transversion changes codon 249 of *P53* from AGG (encoding arginine, a basic amino acid) to AGT (encoding serine, a small nucleophilic amino acid).

The mutagenic properties of AFB1 are acquired upon its metabolic conversion to its exo-8,9-epoxide form. The AFB1-epoxide reacts directly with guanine and forms a number of distinct adducts. These adducts are chemically reactive and promote depurination of the G and ultimate replacement of the original G with the pyrimidine T. The formation of adducts appears to be favored at the second G in GG dinucleotides, with the modification and subsequent mutation occurring at the second G. The base 3′ to the modified G also seems to confer some degree of site specificity. Overall, the known sequence biases do not fully account for all of the hotspots at which AFB1 has been shown to act, indicating that some additional structural factors remain to be discovered.

Inflammation Promotes the Propagation of Cancer Genes

As we have seen from the preceding examples, environmental carcinogens can directly convert normal genes to cancer genes by inducing mutations. In addition to their direct effects on DNA sequences, carcinogens can also promote the development of cancer by promoting the growth of cells that have acquired mutations. Most well-defined carcinogens induce the creation of a *microenvironment* in which mutations are more likely to occur, and in which cells that harbor cancer genes can preferentially proliferate. This microenvironment is created by the inflammatory response.

Inflammation is both a risk factor for initial cancer development as well as a consistent component of the microenvironment of established cancers. The relationship between inflammation and cancer was recognized as early as 1863 by Rudolf Virchow,

who noted that some types of irritants could enhance cell proliferation. We now understand that increased cell proliferation alone does not cause cancer. Rather, inflammation simultaneously produces mutations and creates an environment where mutated cells will tend to proliferate.

The dual effects of carcinogens in the generation of mutations and in the subsequent proliferation of mutant cells are well illustrated by asbestos. Exposure to asbestos is a strong risk factor for the development of mesothelioma, a relatively rare cancer that affects the lining of the lungs and the pleural cavity.

Environmental asbestos occurs in a number of fibrous forms that each have an integral iron component. The physical properties of asbestos fibers made them a widely used component of fireproof ceramics and insulation until the association of asbestos and lung disease was appreciated. It appears that these physical properties, combined with intrinsic chemical reactivity, make asbestos a potent carcinogen. Ingested by inhalation, asbestos fibers are engulfed by cells of the immune system by the process of phagocytosis. Longer fibers are incompletely phagocytized and are inefficiently cleared from the lungs. Asbestos fibers are essentially a chronic irritant that triggers a strong inflammatory response, known as asbestosis.

The presence of asbestos in the lung leads to recruitment and activation of inflammatory cells, including pulmonary alveolar macrophages and neutrophils. The mediators of asbestos toxicity are reactive oxygen species and reactive nitrogen species, which, as we have seen previously, can damage DNA. Reactive oxygen species, including superoxide radicals and hydrogen peroxide, and reactive nitric oxide are released by activated inflammatory cells and irritated parenchymal cells. In addition, it has been shown that free radicals can be directly generated by asbestos fibers in cell-free systems, a reaction thought to be directly catalyzed by the iron component. Thus, there are two distinct sources of potentially mutagenic reactive species: the cells that are irritated by the asbestos fibers, and the fibers themselves.

Chronic inflammation is an important predisposing factor for many human cancers. It is estimated that chronic inflammation contributes to approximately one quarter of all malignancies. The best evidence that supports a role for inflammation in tumorigenesis is the clear relationship of inflammatory diseases and cancers. Diseases that have a significant inflammatory component can strongly predispose affected individuals to cancer. Some inflammatory diseases, like asbestosis, are related to an environmental exposure, while the etiology of others is less well understood. Among the strongest links between chronic inflammation and carcinogenesis is the association between the inflammatory bowel diseases ulcerative colitis and Crohn's disease with the development of colon cancer. Chronic inflammation has also been shown to be a significant risk factor for cancers of the esophagus, stomach, liver, prostate and urinary bladder. The etiology of the inflammation varies in these diseases but the relationship between chronic inflammation and the later development of cancer is similar.

Infectious agents are significant cause of chronic inflammation that gives rise to cancer. Accordingly, infectious agents that cause chronic inflammation have been shown to increase cancer risk. Collectively, infectious agents are thought to

contribute to approximately 15% of all cancers worldwide. Virus-associated cancers are particularly common and represent a significant, but theoretically tractable, public health problem.

The relationship between viruses and cancer is complex and largely beyond the scope of this text. Several carcinogenic viruses integrate into the genome and alter endogenous genes or deliver viral genes. The human papillomaviruses affect the epithelial cells of the uterine cervix by the transfer of genetic material (see Chapter 6). Another example is Herpesvirus 8, which integrates into the precursor cells of Kaposi sarcoma. Aside from these two cancers, most available evidence suggests that viruses and other infectious agents most often contribute to cancer indirectly by inducing host inflammatory responses.

The Hepatitis B and C viruses cause chronic inflammation of the liver and facilitate the subsequent development of liver cancer. In parts of Asia, the combined effects of Hepatitis virus infection and exposure to the mutagen aflatoxin B1 cause a 1,000-fold increase in cancer risk. Numerous infectious agents that cause chronic inflammation and significantly increase cancer risk (see Table 1.2).

How does inflammation contribute to the development of cancer? The relationship between these two complex entities remains to be completely understood, but several aspects are clear. One contributing factor is the creation of somatic mutations by free radicals. As we have seen in the case of the potent carcinogen asbestos, free radicals can be generated by both the agent and by the cellular component of the immune response. Infectious agents typically induce a strong cellular immune response, which leads directly to a free radical response. Leukocytes and other phagocytic cells normally produce these highly reactive species to kill and denature infectious agents. Reactive oxygen and nitrogen species react to form peroxynitrite, a powerful mutagen. Mutagenesis therefore appears to be a byproduct of a vigorous immune response.

Another important factor in cancer development is the humoral component of the inflammatory response: the local production of signaling proteins known as cytokines and chemokines. These molecules are potent stimulators of cell division and function to recruit additional immune cells and activate local fibroblasts that

Table 1.2 Chronic inflammation and cancer predisposition. Many cancers are preceded by a local inflammatory response to an infectious agent

Infectious agent	Type	Inflammatory disease	Cancer
Hepatitis B virus Hepatitis C virus	DNA virus	Hepatitis	Liver cancer
Helicobacter pylori	Bacterium	Gastritis	Stomach cancer
Epstein–Barr virus	DNA virus	Mononucleosis	B-cell, non-Hodgkin's lymphoma Burkitts lymphoma
Human Papillomavirus	DNA virus	Cervicitis	Cervical cancer
Schistosoma haematobium	Trematode	Cystitis	Bladder cancer
Opisthorchis viverrini	Flatworm	Cholangitis	Bile duct cancer

will amplify the inflammatory response. Also secreted by activated cells are proteolytic enzymes that break down the extracellular matrix and thereby alter the tissue structure. These changes can alter cell spacing and make cells more mobile. Finally, the secretion of angiogenic peptides promotes the growth of new vasculature, which expands the spaces where cells can thrive. The combined affects of these changes appear to make a fertile environment for the proliferation of cells that have acquired cancer genes and the subsequent growth of tumors. The humoral component of inflammation thus changes the microenvironment to favor the proliferation of cells with cancer genes.

Inflammation can play a significant role in two distinct stages of a cancer: tumor initiation and subsequent tumor growth and progression. While it appears that the majority of cancers arise in the absence of a known chronic inflammatory condition, inflammatory cells contribute to the microenvironment of nearly every established tumor. When analyzed histologically, established tumors are typically found to contain large numbers of infiltrating inflammatory cells (see Fig. 1.12). Indeed, a significant proportion of the mass of a typical tumor is comprised of cells produced by the immune system. Viewed histologically and as gross specimens, cancers resemble wounds that do not heal.

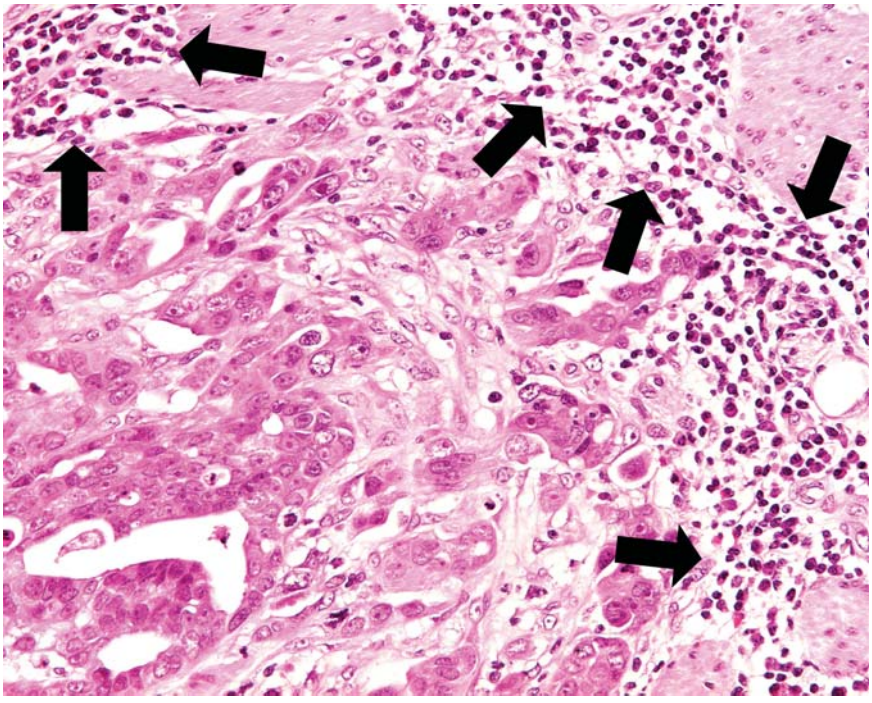

Fig. 1.12 Cancers exhibit areas of chronic inflammation. Inflammatory cells (indicated by arrows) are present throughout this section of a stomach adenocarcinoma. (Courtesy of Angelo De Marzo M.D., Ph.D., Johns Hopkins University.)

It is not difficult to imagine how the profusion of mitogenic stimuli, the weakening of the extracellular matrix and the onset of angiogenesis that occurs in inflamed tissues might promote the continued clonal proliferation of cells with cancer genes. It is important to remember the obvious fact that the function of the immune system is not to promote cancer. On the contrary, the inflammatory response seen in established tumors may be a futile attempt by the host immune system to eliminate those tumors. It is likely that many early tumors do in fact die off in the miasma created by the immune system, which is in many regards toxic. The growth of tumors might be best characterized as an ongoing battle between cancer cells and the immune system. This battle is gradually lost as cancer cells acquire new phenotypes that allow them to survive and proliferate where normal cells would fail to thrive. As we will see, evolutionary theory provides an explanation of why a force mobilized to defeat a cancer might end up promoting it instead.

Darwinian Selection and the Clonal Evolution of Cancers

In the preceding sections, we have seen how mutations arise. It this section and those that follow, we will explore how individual mutations can accumulate in a single cell lineage and give rise to a tumor.

Most neoplasms are believed to arise from a single cell. Several lines of evidence support this idea. In studies conducted prior to the availability of molecular genetic approaches, it was observed that the pattern of X chromosome inactivation is typically uniform in cancer cell populations, which is indicative of a single precursor. Lymphoproliferative neoplasms that produce immunoglobulins almost always produce a single, clonal isotype. Finally, genetic analysis of primary tumors, from the level of DNA sequence to whole chromosomes, typically reveals mutations and structural changes that are present in all tumor cells, suggesting a unicellular origin. The preponderance of evidence indicates that the cancer cells that ultimately compose a tumor mass are vertically derived from a founder cell and therefore contain the same cancer genes. In this sense, individual tumors are monoclonal.

How do these clones arise? When a somatic mutation occurs in a single cell there exists for a time only a single copy of that newly acquired mutant allele. That mutation will become more widespread if that original cell divides and gives rise to progeny that also contain the mutant gene. Put another way, the mutant clone expands by the process of cell proliferation. A somatic mutation that occurs in a non-dividing cell would not expand as a clone and could therefore not contribute to a cancer. Progression of a cancer requires clonal expansion of cells that harbor cancer genes.

Why do cell clones that harbor cancer genes expand? In an elegant hypothesis presented in 1976, Peter Nowell described how cancer genes confer a selective advantage that allows cells to essentially outcompete neighboring cells. This phenomenon is in many ways analogous to speciation as explained by Charles Darwin's theory of evolution. Natural selection occurs when an individual organism occupies

a niche in which that organism's genotype confers an advantage. That advantage is selectable if it promotes the production of more progeny. New niches present new opportunities for individual genotypes to potentially thrive. Advantageous proliferation within a niche can eventually lead to speciation. In several respects, tumorigenesis can be viewed as a form of cellular speciation.

Tissues represent a cellular niche. A region of tissue that encompasses a cellular niche is often called a *compartment*. In adults, the number of cells that occupy a self-renewing tissue compartment, such as the epithelial lining of the gastrointestinal tract or the marrow within bony trabeculae, is normally stable. Stability depends upon the balance between two opposing processes: cell birth and cell death.

The cells that proliferate within a compartment and give rise to the diverse cellular components of a given tissue are known as *stem cells*. Stem cells are both proliferative and immature. In a stable compartment, the number of cells that arises though cell division is equal to the number of cells that mature into individual functionally specialized cells, stop proliferating and ultimately die. Cells thus enter the compartment via the proliferation of stem cells, perform their functions as mature, non-proliferating cells, and exit the compartment via cell death (see Fig. 1.13). Cells that occupy highly proliferative compartments typically possess an intrinsic and highly regulated program that actively induces cell death. This form of programmed cell death is known as *apoptosis*. Apoptosis is distinct from cell death that results from insult or injury in that it contributes to the stability of that tissue compartment.

The fine balance between stem cell proliferation and apoptotic cell death dictates the stability of a given compartment, or what is known as *tissue homeostasis*. Tissue homeostasis is disrupted when the rate of cell birth is unequal to the rate of maturation, cell death and removal.

Cancer genes cause a disruption in tissue homeostasis. If a gene confers a phenotype that increases proliferation or prevents maturation or cell death, then the cells that harbor that gene may begin to outnumber other cells in that compartment and form a neoplasm. This is the first stage of the clonal evolution of a tumor.

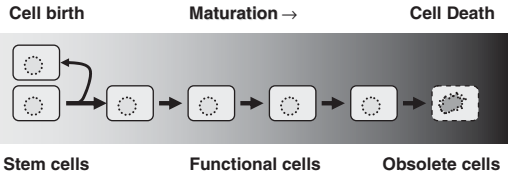

Fig. 1.13 Homeostasis within a tissue compartment. Stem cells undergo an asymmetrical division in which one daughter cell is fated to mature and the other remains an undifferentiated stem cell. Thus stem cell populations self-renew. Mature cells carry out the various functions of the tissue, until they reach the end of their life spans and are eliminated from the compartment. In stable compartments, the rate of cell birth is equal to the rate of cell death. Highly proliferative compartments can be completely renewed in several days

Selective Pressure and Adaptation: Hypoxia and Altered Metabolism

The precise forces that favor the selection of cells which harbor cancer genes remain incompletely understood. The causal relationship between inflammation and cancer provides a significant clue as to the nature of clonal selection and how the acquisition of cancer genes can facilitate adaptation, survival and proliferation.

As described previously, many cancers arise in areas of chronic inflammation. Inflammation creates numerous changes in the microenvironment of a cellular compartment. The activation of free radical-producing cells, the release of humoral factors and secretion of enzymes combine to alter tissue structure and to change oxygen, glucose and pH levels. These changes produce selective pressure. Cells that can continue to proliferate in these conditions would be more likely to survive as a viable clone. In the Darwinian sense, inflammation creates a new niche within a cellular compartment. A key question is: how do cancer cells adapt to new niches?

To illustrate the role of adaptation in clonal evolution, we will consider a single cellular characteristic that changes during tumorigenesis: metabolism. Cancer cells acquire altered metabolic states that enhance survival in adverse microenvironments.

In 1930, Otto Warburg observed that the metabolism of cancer cells differs from that of normal cells. While normal cells produce energy primarily by aerobic respiration, the cells in tumors rely more heavily on glycolysis, an anaerobic reaction. The metabolic switch that occurs during tumorigenesis has subsequently come to be known as the *Warburg effect*.

Glycolysis is relatively inefficient. While 36–38 molecules of ATP are produced by the complete oxidation of one molecule of glucose, only 2 ATP molecules are generated by the anaerobic conversion of glucose into pyruvate.

Glycolysis

$$\text{Glucose} + 2\ P_i + 2\ \text{ADP} + 2\ \text{NAD}^+ \rightarrow$$

$$2\ \text{pyruvate} + 2\ \text{ATP} + 2\ \text{NADH} + 2\ H^+ + 2\ H_2O$$

Oxidative phosphorylation

$$\text{Glucose} + 36\ \text{ADP} + 36\ P_i + 36\ H^+ + 6\ O_2 \rightarrow$$

$$6\ CO_2 + 36\ \text{ATP} + 42\ H_2O$$

Additionally, the hydrogen ions produced as a byproduct of the glycolysis reaction cause the acidification of the cellular microenvironment. Cancer cells thus appear to acquire a phenotype that is both energetically inefficient and environmentally

toxic. The obvious drawbacks of this metabolic switch are apparently outweighed by one critical attribute: the ability to survive oxygen deprivation.

The structure of normal tissues is constrained by blood supply. Blood flow enhances tissue oxygenation and thus facilitates aerobic respiration. Cell proliferation is a process that requires a significant amount of energy – energy that in normal cells is generated via oxidation of glucose. Proliferation of normal cells is therefore favored in regions that are well oxygenated. Conversely, proliferation of normal cells is limited in tissue spaces that have low oxygen tension, an environmental state known as *hypoxia*. In normal cell compartments, proliferation is spatially restricted to regions that are close to the local blood supply. Experimentally, hypoxia has been detected in tumor tissues that are more than 100 microns from the nearest blood vessel.

In areas that are relatively distant from the blood supply, hypoxia creates a niche with a distinct selective pressure. While cancer cells tend to be inefficient and toxic in their metabolism, they also have a lower reliance on oxygen because their glycolytic pathways are upregulated. Cancer cells have thus adapted to a niche that is inhospitable to normal cells. It is possible that the acidification of the microenvironment that occurs as a result of increased glycolosis creates an additional form of selective pressure.

Multiple Somatic Mutations Punctuate Clonal Evolution

Genetic analysis of cancer samples invariably shows that cancer cells contain multiple cancer genes. This implies that multiple somatic mutations are required during the process of tumorigenesis. A large body of experimental evidence has shown that this is in fact the case. How does the process of clonal evolution relate to the acquisition of multiple mutations?

Up to this point, we have seen how a single cancer gene might be acquired. Somatic mutations can occur by a variety of processes, including the stochastic deamination of methylated cytosines, errors during DNA replication and repair, and chemical mutagenesis caused by environmental carcinogens and inflammatory agents. In rare instances, these somatic mutations will change a normal gene into a cancer gene. A cell clone harboring a cancer gene will proliferate if that cancer gene provides a unique advantage that allows it out compete its neighbors that do not harbor the mutation. This outgrowth of cells becomes a microscopic neoplasm.

What happens next? In most cases, nothing happens. In tissues that have been carefully studied, it appears that most neoplastic clones fail to progress and eventually die off. Most neoplasia represent a dead end for that clonal lineage. In these cases, the growth advantage attained by a neoplasm is apparently not sufficient to allow sustained expansion. Perhaps the expanding cell clone encountered a new selective pressure, such as a successful immune response by the host. An expanding cell clone might also fall victim to the byproducts of its own proliferative success by contributing to a critical shortage of oxygen or overabundance of metabolically

derived acid. The barriers to tumor growth, and therefore the selective pressures that appear as a tumor grows, are likely to vary significantly in different tissues.

Proliferating cell clones are neoplasia by definition, but not all neoplasia develop into cancers. Just as a very small proportion of cells that are mutated give rise to neoplasia, only a small proportion of neoplasia progress to cancer. Again, this is directly analogous to the evolution of biological life forms. Most genetic changes are predicted to lead to either no advantage or a disadvantage. In biology as in the biological microcosm that is cancer, only the rare mutation creates a selective advantage. The neoplasia that do progress to tumors are indeed rare products of clonal evolution.

It is reasonable to assume that as a neoplasm grows, the local microenvironment undergoes changes. Concomitant with cell proliferation is the local decrease in the concentrations of metabolic precursors and an increase in metabolic products. As the number of cells increase, the ratio of the cells that occupy the periphery of the neoplasm (which contact the neighboring normal cells) to the cells that are in the middle of the neoplasm (which only contact other cells of the proliferating clone) gets progressively smaller. The space occupied by the proliferating cell mass will alter the spacing between adjacent cells and one another and between all cells and the nearest blood vessel. Local oxygen, glucose and hydrogen ion concentrations will all change. Once a tumor becomes invasive, the cells at the leading edge of the invasion encounter new niches with unique barriers. Finally, the cells that break free of the original tumor mass and metastatize to distant parts of the body will survive detachment, transit through the blood or lymphatic system, and reseeding to grow a new tumor, often in a different type of tissue.

During each stage of tumorigenesis, newly acquired genetic alterations confer new properties to the tumor cells. The rare neoplasia that progress acquire additional cancer genes by somatic mutation. In such cases, a clone containing the initiating mutation or mutations expands and eventually a single cell within that clone acquires an additional mutation that confers an additional growth advantage. This cell gives rise to a new clone that is better adapted to growth in the contemporary microenvironment. The new clone outgrows the previous clone and continues to expand. In this manner, multiple rounds of mutation followed by waves of clonal expansion eventually give rise to a cancer (see Fig. 1.14).

Clonal evolution is an iterative as well as a dynamic process. The two steps of this process are somatic mutation and clonal expansion into constantly changing niches. Both steps are equally important. Somatic mutation gives rise to the phenotypes that favor improved growth and survival, while clonal expansion provides cellular targets for mutation.

How Many Mutations Contribute to a Cancer?

By current estimates, the human genome contains 20,000–25,000 protein-coding genes. About 350 genes – more than 1% of the total – have been found to be mutated in multiple cancers, and are therefore probable cancer genes (see appendix).

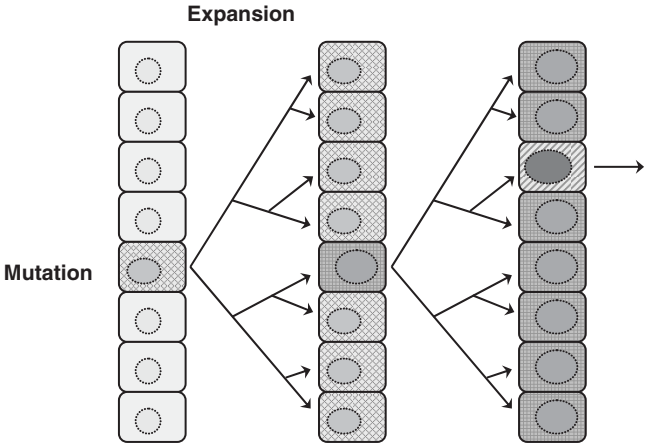

Fig. 1.14 Clonal evolution of tumor cells. A single cell in a normal tissue acquires an alteration that confers a growth advantage. That cell divides and thus expands over time into a distinct clone. A cell within that clone acquires a second mutation that provides an additional growth advantage. A tumor results from iterative rounds of mutation and clonal expansion. (Concept from *The Genetic Basis of Human Cancer* Kinzler and Vogelstein, eds., McGraw Hill (2002).)

Of these, approximately 90% are somatically mutated in cancers, 20% have germline mutations that predispose to cancer and 10% exhibit both somatic and germline mutations.

Mutations that occur in cancers fall into two functional categories: (1) mutations that are required for tumorigenesis; and (2) mutations that merely occur during tumorigenesis and do not contribute to the process. These mutations have been aptly referred to as *drivers* and *passengers*, respectively. Mutations that create cancer genes are drivers, by definition. Drivers confer selective advantages during clonal evolution, and thus 'drive' the process forward. In contrast, passenger mutations do not appear in tumors as a result of evolutionary selection. Rather, a passenger mutation occurs by chance in a cell that harbors a driver mutation. As a clone that contains a cancer gene expands, the passenger mutation merely comes along for the ride.

Recent studies have revealed the first detailed look at the cancer genome. In groundbreaking studies undertaken at Johns Hopkins University in the USA and at the Sanger Institute in the UK, hundreds of protein coding regions were examined in numerous cancer specimens by extensive DNA sequencing. These high-throughput strategies yielded a greater diversity of cancer-associated mutations than had been anticipated. Roughly 100 genes were found to be mutated in each advanced colorectal or breast cancer that was examined in detail. Of these, at least 15–25 were estimated to be driver mutations, with the remainder representing either passenger mutations or genes that were selected at a rate that was below the statistical threshold for significance.

Some cancer genes are significantly more prevalent than others (see Chapter 6). The mutations found within any two cancers are typically different (see Fig. 1.15).

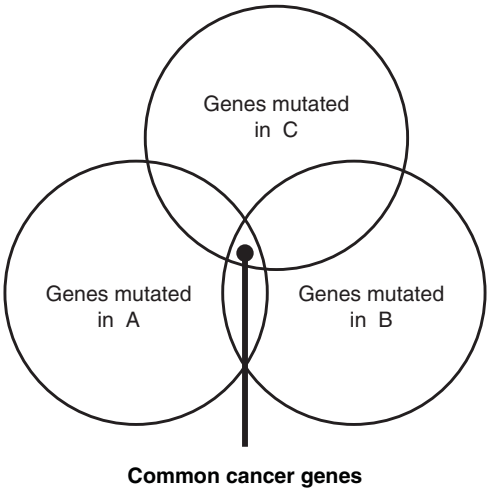

Fig. 1.15 Common and unique drivers of breast cancer. The driver mutations found in breast cancers are diverse and largely tumor-specific. In this example, comparison of the set of genes mutated in tumors A, B and C showns that most driver mutations are unique to each tumor in which they occur, but there is also significant overlap. A small proportion of mutated genes are common to all three tumors; these are likely to represent highly prevalent cancer genes.

However, in many cases mutations in critical cancer genes are found in a large proportion of cancers of a given type. Such mutations can point to cellular processes that are typically defective in a particular cancer type (see Chapter 5).

High-throughput approaches have detected mutations in over 1,000 different genes in two common types of cancer: colorectal and breast cancer. Among these mutant genes, a significant number have been found to also be present in diverse tumor types, in addition to breast and colorectal cancers.

The 15–25 genetic alterations that drive breast and colorectal cancers are a significantly greater number than previous approaches had predicted. Efforts to recapitulate the cancer phenotype *in vitro* have yielded smaller estimates. For example, Robert Weinberg and colleagues showed that the experimental introduction of as few as four genes could alter the properties of cultured normal human cells so that they were able to form tumors when injected into mice. A comparison between the results of such experiments and the subsequently determined number of cancer genes found in actual tumors suggests that the *in vitro* experiments informed a model that was overly simplified. In retrospect, a cell culture vessel is unlikely to fully recreate the complex microenvironment in which tumors grow and in which clonal selection takes place. As will be described in later sections, *in vitro* experiments have nonetheless been very useful in identifying cancer genes and in explaining how some cancer genes are likely to work.

Another approach to quantify the extent of gene mutation in cancers is to sequence individual candidate cancer genes. Candidate cancer genes are those that have been chosen for mutational analysis by virtue of: (1) their transmission in

cancer-prone families, (2) their presence near known areas of chromosomal abnormalities in cancers, (3) their known cellular functions, or (4) their relatedness to genes of known cellular function. The candidate-gene approach has been highly successful. Several of the most prevalent cancer genes were discovered in this manner. The inherent bias in this approach precluded the discovery of cancer genes that are mutated in fewer cancers, but which still contribute significantly to the process of tumorigenesis. Also undercounted by this approach were genes that have a modest effect on readily measurable cellular properties such as cell proliferation and survival, and which were therefore not included on candidate lists. Such studies had resulted in estimates in the range of 5–7 cancer genes per cancer, significantly less than what has actually been found.

Although the genomes of the majority of cancer types remain to be studied in detail, different types of cancers appear to have different numbers of cancer genes. This is probably related to the fact that different tissues have different intrinsic barriers against clonal expansion. For example, one might expect that the liquid tumors, the leukemias and the lymphomas, require fewer cancer genes than breast and colorectal cancers because of the relative lack of physical barriers that prevent their spread.

This hypothesis is supported by epidemiological evidence. The most common cancers are diseases that primarily afflict older individuals. The incidence of carcinomas dramatically increases with age, with a 100-fold increase in incidence occurring over an average lifetime (see Fig. 1.16). The clonal evolution of the common cancers occurs in a time frame that is most often measured in decades. The final expansion of cancer clones with 15–25 mutations is mostly seen in the aged. Conversely, the most common malignancies in young patients are leukemias, which require fewer alterations.

Highly informative data has been derived from studies of the survivors of the atomic bombs in Hiroshima and Nagasaki, who have been closely followed since the end of the Second World War. Leukemias directly attributed to the high levels of

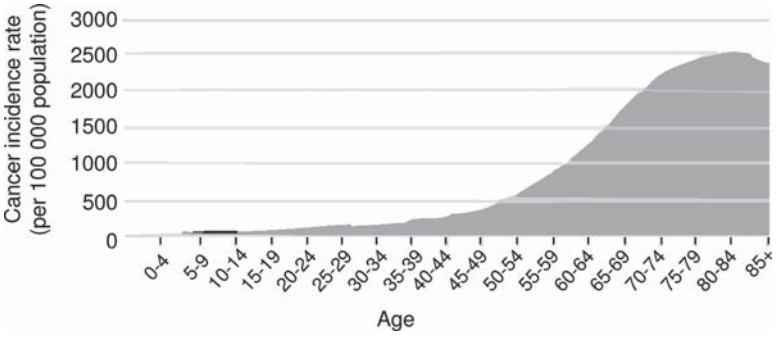

Fig. 1.16 Cancer incidence is age-dependent. The overall incidence rate of cancer dramatically increases in the older age groups. Shown are combined data from all sites and both sexes. (Data from NCI SEER program 1994–1998.)

ionizing radiation associated with the explosions began to appear within three years and the incidence of these cancers peaked by seven years. The increased incidence of solid tumors was not evident until more than ten years after the initial exposure.

Colorectal Cancer: A Model for Understanding the Process of Tumorigenesis

Clonal evolution is an interesting hypothesis that incorporates the concepts of mutation, clonal expansion and population dynamics to explain how tumors arise in cellular compartments. But is it real? What is the evidence that tumors actually arise in the stepwise manner consistent with clonal evolution?

The best evidence comes from exhaustive studies of tumors in the large bowel. Tumors that arise in the epithelium of the colon and rectum are very common. Nearly one half of the US population is affected by colorectal tumors, most of which are benign. Approximately 5% of the population will develop colorectal cancer, the second leading cause of cancer death. The most common histological type is the adenocarcinoma that arises from epithelial cells.

Unlike many types of tumors, tumors of the colon and rectum are highly accessible. Through the use of endoscopy, a widely used screening technique, colorectal tumors can be directly visualized at all different stages of growth and dissemination. At the time of diagnosis, tissue specimens can be readily obtained for the purpose of DNA analysis. The high prevalence and accessibility of colorectal tumors have provided a unique opportunity to study the genes that contribute to tumorigenesis. Collectively, these studies have provided a paradigm for understanding how the accumulation of cancer genes gives rise to a cancer.

The gastrointestinal system is composed of readily defined tissue compartments. Several cell types contribute to the luminal surface of the gastrointestinal tract known as the mucosa (see Fig. 1.17). The normal mucosal surface of the colon is composed of invaginations known as crypts, which function to maximize the surface area of the large bowel. These crypts are lined with a single layer of epithelial cells of three different types: absorptive cells, mucus-secreting goblet cells, and neuroepithelial cells. At the base of each crypt are 4–6 stem cells, which give rise to the mature cells of the crypt. Cells predominantly multiply in the lower one third of the crypt, differentiate in the upper two thirds and are eventually extruded at the apex of the crypt and thereby lost into the lumen (see Fig. 1.18). The epithelial cells of a crypt are a clonal population derived from a self-renewing population of stem cells. Colonic crypts are thus a well-defined cellular compartment, where cells are born, mature, function and die in a linear space.

The smallest colorectal neoplasm that is observable within the colonic mucusa, either by microscopy or by staining with the dye methylene blue, is the *aberrant crypt focus* (ACF). These lesions can affect one crypt or span several adjacent crypts. An ACF is the earliest indication that the delicate balance between cell birth, maturation and death within a crypt has been perturbed.

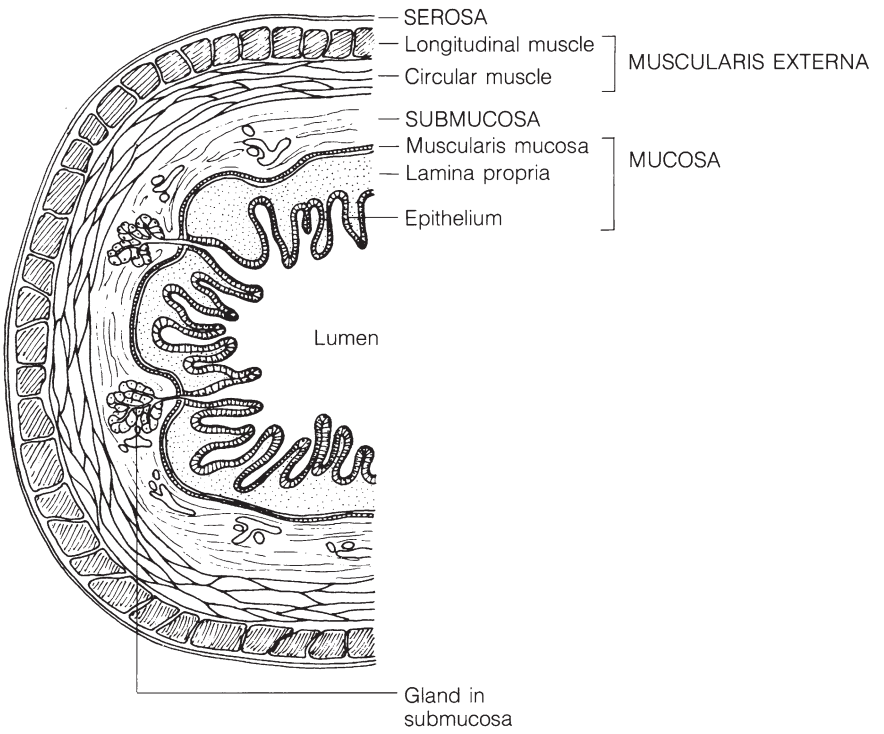

Fig. 1.17 The lining of the gastrointestinal tract. The innermost layer is the mucosa, a membrane that forms a continuous lining of the entire gastrointestinal tract. In the large bowel, this tissue contains cells that produce mucus to lubricate and protect the smooth inner surface of the bowel wall. Connective tissue and muscle separate the muscosa from the second layer, the submucosa, which contains blood vessels, lymph vessels, nerves and mucus-producing glands. Next to the submucosa is the muscularis externa, consisting of two layers of muscle fibers – one that runs lengthwise and one that encircles the bowel. The fourth layer, the serosa, is a thin membrane that produces fluid to lubricate the outer surface of the bowel so that it can slide against adjacent organs. (Courtesy of the National Cancer Institute.)

The earliest readily observable manifestation of a colorectal tumor is the polyp, a growth of cells that often extends into the bowel wall and projects into the intestinal lumen (see Fig. 1.19). Polyps fall into two histological classes: non-dysplastic (also called hyperplastic) and dysplastic (also called adenomatous) polyps. Non-dysplastic polyps have an ordered epithelial structure that is similar to that of normal crypts. These tumors are benign and are thought to have a low tendency to progress. In contrast, adenomatous polyps exhibit a significant degree of histologically apparent dysplasia (see Fig. 1.20). Epithelial cells can line up in multiple layers, and frequently have enlarged nuclei at atypical locations within the cell. Larger adenomas often contain projections of dysplastic crypts that confer what is known as a ‘villous’ morphology. Adenomas become more dysplastic as they grow larger in size. With size they also become more likely to invade surrounding tissues, at which point they are defined as malignant.

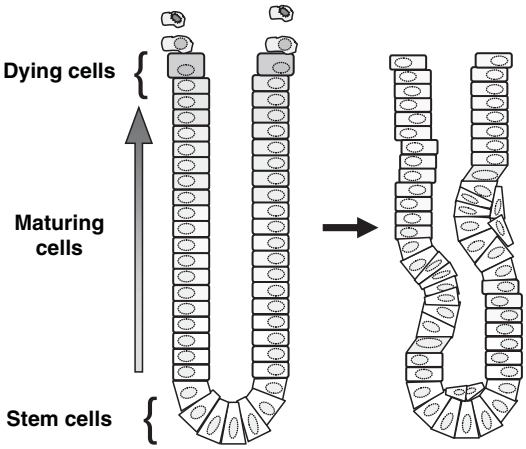

Fig. 1.18 Cell birth and death in a colon crypt. The invaginations of the colorectum form structurally defined tissue compartments known as crypts. In this simplified representation, an increase in cell birth or decrease in cell death leads to hypercellularity and loss of tissue organization.

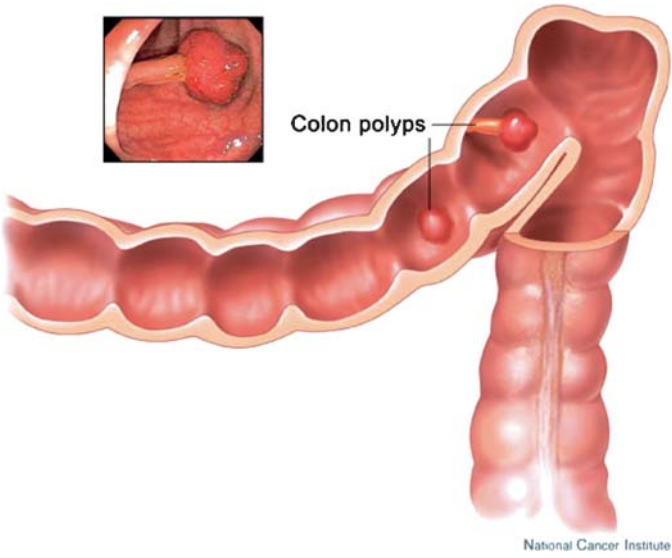

Fig. 1.19 Colon polyps. Polyps are tumors within the colorectal mucosae. Two colon polyps, one flat and one pedunculated are shown. Inset shows photo of a pedunculated polyp. (Illustration by Terese Winslow, courtesy of the National Cancer Institute.)

Tumor growth can be discontinuous. For example, a small polyp may remain dormant for years or even decades. But when a subsequent mutation occurs in one cell, a new wave of expansion can occur. A significant proportion of adenomas progress and become malignant tumors. Size is a reliable indicator of malignant

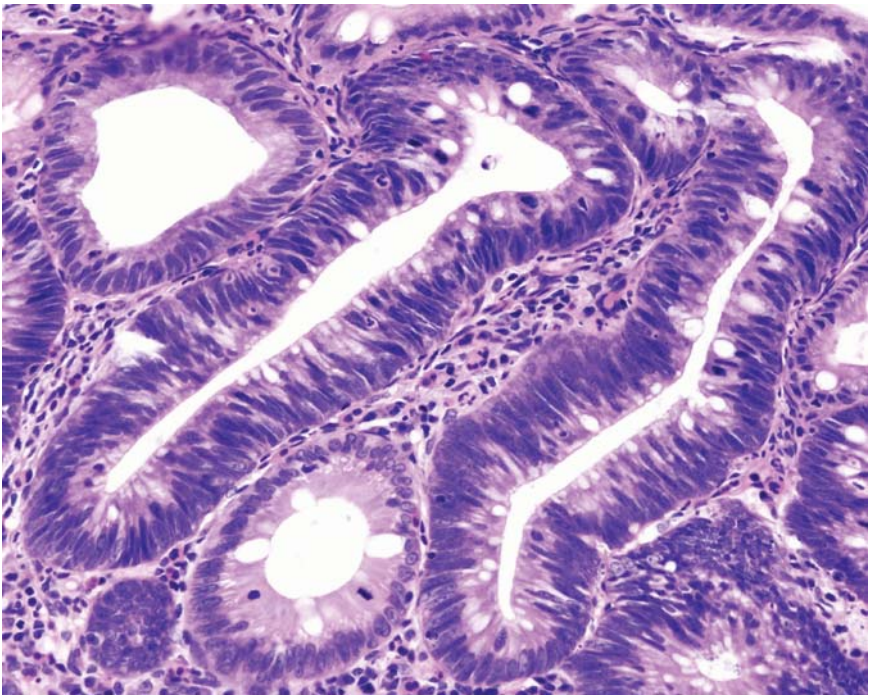

Fig. 1.20 Histology of an adenomatous polyp. A section of an adenoma removed during endoscopy, stained with hematoxylin-eosin, shows mild dysplasia.

potential. Few adenomas that are less than 10 mm in diameter will progress into a malignancy, but adenomas larger than 10 mm will have an estimated 15% chance of becoming malignant in the subsequent 10 years. Advanced colorectal tumors can locally metastasize to the mesenteric lymph nodes (see Fig. 1.21), or travel more distantly, typically to the peritoneum and the liver.

Benign polyps can usually be resected during colonoscopy, while malignant tumors require more extensive surgery for their excision. The probability of a cure is significantly lower if a tumor has metastasized. In such cases, surgery is combined with a form of adjuvant therapy such as chemotherapy or treatment with ionizing radiation. While such treatments can achieve remission, about 40% of these patients will die from their disease within 5 years of the initial diagnosis.

In seminal studies conducted in the 1980s and 1990s, Bert Vogelstein, Kenneth Kinzler and their coworkers demonstrated how genetic alterations underlie the progression of colorectal tumors. The illustration of the defined stages of a colorectal tumor combined with the gene changes commonly associated with these transitions is informally referred to as a Vogelgram (see Fig. 1.22). These studies form a paradigm for understanding how multiple cancer genes contribute to tumorigenesis. Both the nature of the mutations and the order in which they are acquired are critical features of this model.

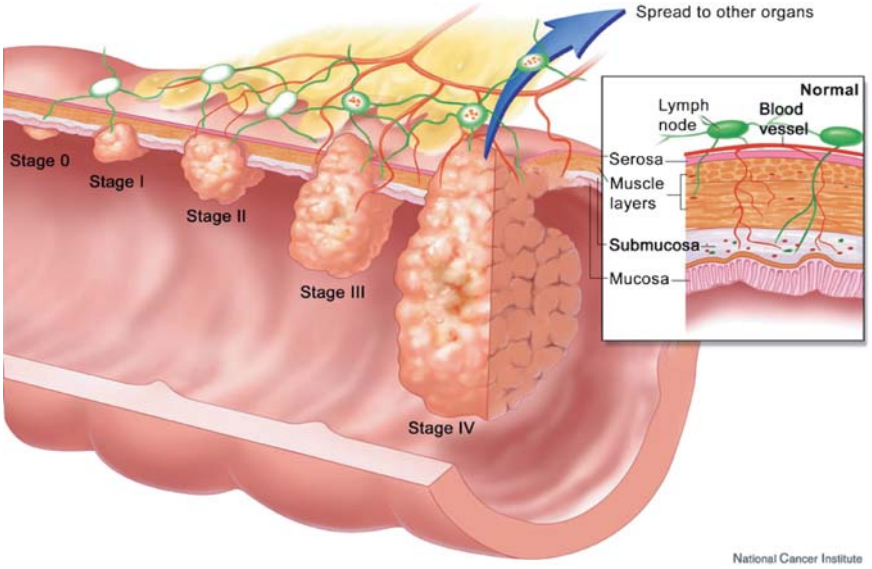

Fig. 1.21 Progressive growth of colorectal tumors. Early-stage tumors are confined to the mucosa. Growing tumors progressively invade the submucosa and muscular layers of the bowel, eventually penetrating the mesenteric vasculature (red) and lymphatic ducts (green). (Illustration by Terese Winslow, courtesy of the National Cancer Institute.)

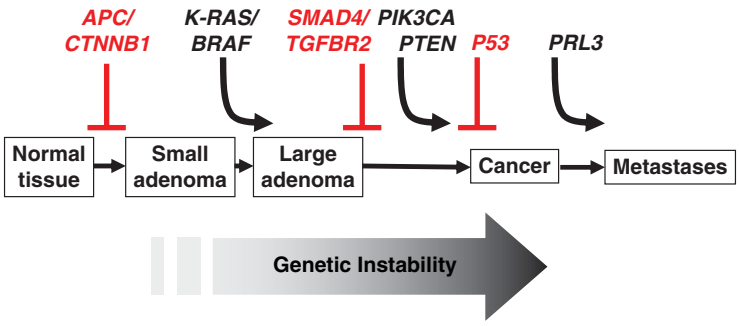

Fig. 1.22 Genetic alterations drive colorectal tumorigenesis. The Vogelgram illustrates the relationship between the histological stages of cancer development and the cancer genes that facilitate clonal expansion. Some cancer genes directly promote the growth of tumor cells (Chapter 2). Other cancer genes remove barriers to tumor growth (shown in red; Chapter 3). The acquisition of successive genetic alterations is accelerated by the process of genetic instability (Chapter 4). Cancer genes combine to affect virtually every aspect of tumor cell growth and death (Chapter 5), in every type of cancer (Chapter 6). Cancer genes are the cause of cancer, but can also lead the way to new treatments (Chapter 7). (Concept from Fearon and Vogelstein, *Cell* 61:759 (1990).)

Do Cancer Cells Divide More Rapidly than Normal Cells?

Cancer is often described in lay terms as a disease caused by cells that are 'out of control'. This is undoubtedly an accurate assessment. Cancer cells do not respond appropriately to the controls that inhibit growth, including spatial, humoral and metabolic signals that would halt the proliferation of normal cells. However, it might be inferred that 'out of control' cancer cells are dividing more rapidly and thus have a shorter doubling time than normal proliferating cells. There is in fact little evidence that this is the case, and several major pieces of evidence that suggest that the opposite may be true.

Malignant tumors such as those that arise in the colorectum are found mostly in older people and result from several decades of clonal evolution. By the time an adenoma reaches 10 mm in diameter, and therefore has the potential to progress into a malignant cancer, it may contain roughly 10^9 cells. This number of cells could theoretically be achieved by only 30 sequential population doublings ($2^{30} = 10^9$), if all progeny continue to proliferate. The proportion of cells within a tumor that can give rise to tumorigenic progeny is a point of ongoing debate. Nonetheless, there is little evidence that growing cancer cells divide at a faster rate than do the stem cells in the base of a normal crypt. The epithelial cells in a normal crypt are replaced every 3–4 days by the proliferation of stem cells at the base of the crypt. At this rate, normal crypt epithelia turn over about 100 times every year. By this simple measure, the stem cells that give rise to normal colonic epithelia appear to be much more highly proliferative in nature than the proliferative tumor cells of colorectal cancers.

An abnormally high proliferation rate is not required to account for a typical tumor mass, given the time frame in which tumors are known to arise. Therefore, from a theoretical perspective, there is no reason to expect that the cells in a neoplasm will proliferate more rapidly than normal dividing cells. The idea behind the clonal evolution model is that neoplasms continue to divide, and fail to die, in changing microenvironments to which they are well adapted. A shorter doubling time would not necessarily confer an additional survival advantage, especially in niches where resources are limiting.

Some cancer phenotypes may in fact impede growth. In many cancers, it is clear that the process of cell division is complicated by chromosome abnormalities. As will be extensively discussed in later chapters, many cancers have abnormal numbers of chromosomes, as well as chromosomal structural abnormalities. These abnormalities are associated with defects in the cellular machinery that monitors the segregation of chromosomes during mitosis. Examination of dividing cancer cells occasionally reveals chromosomes trapped between two separating daughter cells, a phenomenon known as an anaphase bridge (see Fig. 1.23). Such defects present a challenge to cell division, and could theoretically make the process of cell proliferation *less* efficient in cancer cells.

In much the same way that cancer cells use an inefficient form of metabolism that allows them to adapt to adverse niches, the cancer cell cycle appears to have

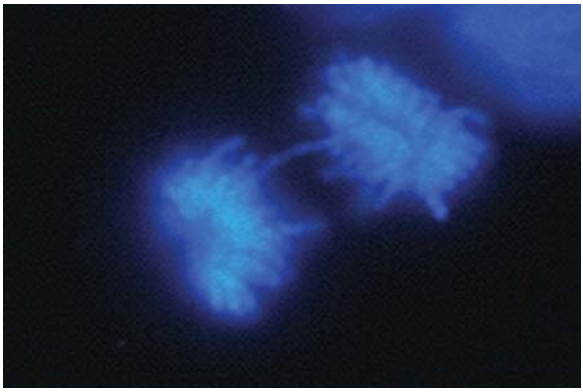

Fig. 1.23 An anaphase bridge. At the end of mitosis, two cancer cells remain connected by incompletely segregated chromosomes. (Courtesy of Dominique Broccoli, Ph.D., Memorial Health University Medical Center.)

defects as well. Presumably, any inefficiency in cell division is outweighed by the evolutionarily benefits conferred by a low level of genetic instability, as will be described in Chapter 4.

Germline Cancer Genes Allow Neoplasia to Bypass Steps in Clonal Evolution

To this point we have exclusively considered somatic mutations as the source of selectable genetic variation. While the majority of cancer genes that contribute to tumor progression are indeed acquired somatically, inherited cancer genes also play an important – and highly illuminating – role in the clonal evolution of some cancers.

Inherited cancer genes can increase cancer risk. A key observation – that ultimately leads to the explanation of cancer predisposition – is that cancers with a strong familial component often occur earlier in life. Clearly, inherited cancer genes must contribute in a significant way to the clonal evolution of tumors.

Well-characterized cancer genes typically exhibit an autosomal dominant pattern of inheritance. In these cases, the presence of only a single allele of a cancer gene causes the associated phenotype, an increased cancer risk. By the laws of Mendelian inheritance, one half of the offspring of an individual that carries such a cancer gene would be expected to inherit that gene and to experience a similarly elevated cancer risk (see Fig. 1.24). It is important to emphasize that while cancer development is dependent on the acquisition of a finite number of distinct, somatically acquired mutations, which cannot be predicted by the laws of Mendel, an increased *risk* of cancer can be transmitted from generation to generation in a Mendelian fashion.

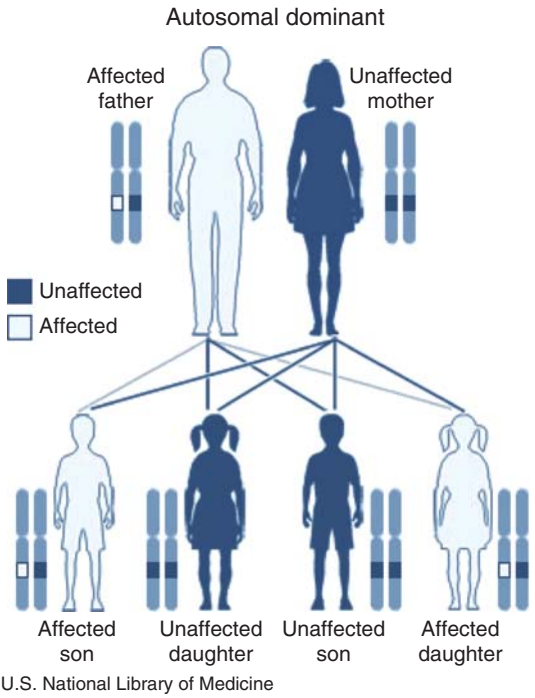

Fig. 1.24 Cancer predisposition can be inherited in autosomal dominant fashion. By this mode of inheritance, one half of the offspring will harbor a germline cancer allele from the affected (cancer-predisposed) parent. (Courtesy of the US National Library of Medicine.)

Germline cancer genes increase the risk of cancer because such genes essentially 'short circuit' the process of clonal evolution that drives the process of tumorigenesis. A germline cancer gene is, by definition, present in every cell of an individual. Therefore, such a gene will be present in every neoplasm that arises. We have seen how the process of tumorigenesis results in the clonal accumulation of multiple mutations. The acquisition of some of these mutations are rate-limiting. That is, a tumor will not be able to progress beyond a certain point without acquiring a particular cancer gene.

A cancer gene that is already in the germline does not have to be reacquired by somatic mutation. In this case, a rate-limiting step in the process of tumorigenesis is eliminated. The presence of a germline cancer gene in an expanding cell clone essentially allows that clone to skip one iteration of mutation and clonal expansion. An inherited cancer gene that circumvents a rate limiting step in tumorigenesis would be expected to increase the overall lifetime risk of cancer and also to cause cancers to arise at a younger age. These observations are entirely consistent with – and thus serve to reinforce – the idea that clonal evolution selects for cells that harbor cancer genes.

Cancer Syndromes Reveal Rate-limiting Steps in Tumorigenesis

The contribution of inherited predispositions to the overall incidence of colorectal cancers has been difficult to ascertain, but estimates of the proportion of colorectal cancers that can be attributed to the inheritance of cancer genes have ranged between 15% and 50%. The genetic basis for most heritable predisposition is unknown. Only 3–5% of all colorectal cancers occur in individuals in well-described syndromes in which the underlying mutations are well described. The majority of colorectal cancers arise in the absence of significant inherited predisposition and are known as *sporadic* cancers. Nonetheless, inherited colorectal cancer syndromes provide important insights into the genetic basis of tumorigenesis.

Colorectal cancers provide a useful model for understanding how the accumulation of somatic mutations leads to the initiation and progression of sporadic tumors. Studies of heritable colorectal cancer syndromes confirm and expand this model by showing how germline cancer genes combine with somatic mutations to short circuit and thereby accelerate the process of tumorigenesis. The contribution of germline cancer genes is exemplified by two heritable colorectal cancer syndromes: *familial adenomatous polyposis* (FAP) and *hereditary nonpolyposis colorectal cancer* (HNPCC). Each of these diseases is caused by the inheritance of a cancer gene in an autosomal dominant manner. Although both of these syndromes predispose affected individuals to colorectal cancer, the effects of these genes on colorectal epithelial cells are very different.

Patients with FAP develop large numbers of colorectal polyps at a young age. Typically, hundreds to thousands of these benign lesions will develop during the second and third decade of life. About one in every 10^6 colorectal epithelial stem cells gives rise to a polyp in these patients. The vast majority of stem cells proliferate normally, and only a small proportion go on to form a observable neoplasm. As in the case of polyps that occur sporadically, the majority of polyps in FAP patients do not progress. However, the sheer number of polyps that arise leads to a significant risk that some of these tumors will progress to invasive, malignant cancers.

The genetic defect in FAP patients is a germline mutation in the *adenomatous polyposis coli* (*APC*) gene that is present in about 1 in 5,000–10,000 individuals. Patients with germline mutations in *APC* have a much greater risk of developing colorectal cancer than the general population, and also often develop manifestations in other tissues including retinal bone and skin lesions and brain tumors. The genetics of *APC* mutations will be described in Chapter 3; the roles of the APC protein in the developing cancer cell will be discussed in Chapter 5.

That the inheritance of a mutant *APC* gene causes a plethora of early colorectal tumors is strong evidence that *APC* mutation affects a rate limiting step in tumor initiation. FAP patients are remarkable because of the number of colorectal tumors that develop. In contrast, the process by which these benign lesions subsequently progress appears indistinguishable from that seen in sporadic tumors. For this reason, *APC* has been described as a *gatekeeper* that is required for maintaining tissue homeostasis. Gatekeepers such as *APC* function in stem cells to keep the proper balance between

cell proliferation, differentiation and death. By this analogy, *APC* mutation opens the gate to the subsequent accumulation of mutations that eventually lead to a cancer.

The role of *APC* mutation in cancer is not limited to the inherited alleles that cause FAP. On the contrary, somatically acquired mutations of *APC* are present in the overwhelming majority of all colorectal neoplasms, most of which are sporadic. *APC* inactivation by mutation is therefore a nearly universal step in the initiation of colorectal tumors. Although FAP is a rare syndrome affecting less than 1% of all families, the genetic analysis of FAP has provided important insights into both the inherited and sporadic forms of a very common cancer.

HNPCC, also known as Lynch syndrome, is another Mendelian disease associated with an increased risk of colorectal cancer. Like FAP, HNPCC accounts for a small proportion of all colorectal cancers that occur in the Western world. The genes that cause HNPCC, and the mechanisms by which mutations in these genes affect disease risk, are clearly distinct from those of FAP. The comparison of these two diseases sheds light on the rate-limiting aspects of colorectal tumorigenesis.

Unlike FAP, HNPCC is not characterized by an increase in polyps. In HNPCC-affected individuals, adenomas occur at the same rate as in the general population. However, the adenomas that do arise in HNPCC patients progress to cancer at an increased rate. These tumors have several unique features. The degree of histological differentiation of these tumors is often low as compared with sporadic tumors of the same size, which normally is an indicator of an aggressive lesion. Contrary to this negative prognostic factor, colorectal cancers in HNPCC patients typically have a better outcome than matched sporadic cancers. This might indicate that HNPCC-associated colorectal tumors evolve somewhat differently than sporadic tumors. HNPCC also affects noncolonic tissues, and affected individuals are at an increased risk of cancers in the endometrial lining of the uterus, small intestine, ovary, stomach, urinary tract, and brain.

While FAP is caused by different mutations within a single gene, *APC*, HNPCC is caused by several different mutant genes that are inherited through the germline of affected families. The genetic heterogeneity of this disease entity complicated epidemiological analysis and obscured the true nature of HNPCC for many years. To this day, the combination of genetic heterogeneity and the high rate of sporadic colorectal cancers in the general population have made the prevalence of HNPCC difficult to quantify.

The genes that cause HNPCC when mutated play a role in the maintenance of DNA replication fidelity. The maintenance of DNA replication fidelity is one of the mechanisms by which the genome is stabilized during multiple rounds of cell division. DNA mismatches that escape the proofreading functions of the replicative DNA polymerases are removed and corrected by a process known as *DNA mismatch repair* (MMR). The genes that are required for this process are mutated in HNPCC. HNPCC thus arises as a result of the failure of the MMR process. Most cases of HNPCC can be attributed to germline mutation of two genes, *hMSH2* and *hMLH1*, with a few cases attributable to a third MMR gene, *hPMS2*. Proteins encoded by these genes function to repair single base pair mismatches and unpaired bases, which tend to occur at high frequency at highly repetitive sequences. Long tracts of

repeat sequences are known as *microsatellites*. The genetic defects that underlie HNPCC tend to cause microsattelite instability, which can be readily measured, and an overall increase in the spontaneous mutation rate. The process of MMR and the contribution of genetic instability to tumorigenesis will be discussed in greater detail in Chapter 4.

The germline cancer genes that cause HNPCC lead to genetic instability and a corresponding increase in the somatic mutation rate. However, HNPCC genes do not appear to contribute significantly to the earliest stages of tumor initiation. The mutation of *APC* initiates the growth of tumors regardless of whether an MMR defect is present or not. Interestingly, the spectrum of *APC* mutations is somewhat different in tumors that exhibit microsattelite instability, suggesting that MMR defects do in fact contribute to *APC* inactivation. The reasons that the MMR defects in HNPCC patients do not lead to an increased rate of *APC* mutation, which would presumably lead to polyposis, is not entirely clear. The increased rate of mutation is instead manifest as an increase in the rate at which the subsequent mutations arise. HNPCC accelerates tumor progression by increasing the rate at which a number of critical somatic mutations are acquired.

Interestingly, both FAP and HNPCC patients develop colorectal cancers at the median age of 42 years, which is 25 years earlier than the median age of patients with sporadic forms of the disease. Given that FAP is a disease of cancer initiation while HNPCC is a disease of tumor progression, the similar age of cancer onset implies that both initiation and progression are similarly rate-limiting.

Understanding Cancer Genetics

In this chapter we have discussed the essential elements of the cancer gene theory. We have seen how cancer genes are acquired and how cancers evolve. These concepts are vividly illustrated by the sporadic and inherited forms of colorectal cancer. The upcoming chapters will delve into the specific genes that cause cancer and how they give rise to the cellular phenotypes that lead to malignancy.

The Vogelgram illustrates several features of the cancer gene theory that explain how sequential genotypic changes cause the evolving phenotypes of growing cancers (see Fig. 1.22). These key concepts will be expanded in the upcoming chapters:

There are two types of cancer genes. Tumorigenesis is driven by mutations that result in the activation of oncogenes (Chapter 2) and the loss of function of tumor suppressor genes (Chapter 3).

Cancers exhibit genetic instability. The rate at which mutations and complex genetic rearrangements occur is not constant during the process of tumorigenesis, but rather increases as genetic alterations accumulate (Chapter 4).

Cancer genes populate intracellular pathways. Cancer genes generally encode proteins that are components of complex molecular circuits, or pathways. In some cases, mutations that disrupt different points in these pathways can similarly trigger clonal expansion (Chapter 5).

Different types of cancers harbor distinct sets of cancer genes. Tumors that arise in different tissues often have characteristic genetic defects in distinct molecular pathways (Chapter 6). These pathways may or may not overlap with those involved in the development of colorectal cancer. Therefore, the Vogelgram describes a process that in detail is specific for cancers that arise in the colonic epithelium, but in principle may be applicable to all cancers.

Cancer genes define potential targets for new forms of therapy. The genes that are altered at different stages of tumorigenesis provide molecular targets for new modes of clinical intervention (Chapter 7). While genes involved in the earlier stages of tumorigenesis might be most useful for cancer prevention and early detection, later mutations highlight potential targets for the treatment of established cancers.

Further Reading

Antonarakis, S. E., Krawczak, M. & Cooper, D. N. Disease-causing mutations in the human genome. *Eur. J. Pediatr*. **159 Suppl 3**, S173–S178 (2000).

Coussens, L. M. & Werb, Z. Inflammation and cancer. *Nature* **420**, 860–867 (2002).

De Marzo, A. M. et al. Inflammation in prostate carcinogenesis. *Nat. Rev. Cancer*. **7**, 256–269 (2007).

Fearnhead, N. S., Wilding, J. L. & Bodmer, W. F. Genetics of colorectal cancer: Hereditary aspects and overview of colorectal tumorigenesis. *Br. Med. Bull*. **64**, 27–43 (2002).

Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759–767 (1990).

Gatenby, R. A. & Vincent, T. L. An evolutionary model of carcinogenesis. *Cancer Res*. **63**, 6212–6220 (2003).

Haber, D. A. & Settleman, J. Cancer: Drivers and passengers. *Nature* **446**, 145–146 (2007).

Hollstein, M. et al. New approaches to understanding p53 gene tumor mutation spectra. *Mutat. Res*. **431**, 199–209 (1999).

Kamp, D. W. & Weitzman, S. A. The molecular basis of asbestos induced lung injury. *Thorax* **54**, 638–652 (1999).

Kelly, P. N., Dakic, A., Adams, J. M., Nutt, S. L. & Strasser, A. Tumor growth need not be driven by rare cancer stem cells. *Science* **317**, 337 (2007).

Klein, C. A. Random mutations, selected mutations: A PIN opens the door to new genetic landscapes. *Proc. Nat. Acad. Sci. USA* **103**, 18033–18034 (2006).

Merlo, L. M., Pepper, J. W., Reid, B. J. & Maley, C. C. Cancer as an evolutionary and ecological process. *Nat. Rev. Cancer* **6**, 924–935 (2006).

Modica-Napolitano, J. S., Kulawiec, M. & Singh, K. K. Mitochondria and human cancer. *Curr. Mol. Med*. **7**, 121–131 (2007).

Nowell, P. C., Rowley, J. D. & Knudson, A. G., Jr. Cancer genetics, cytogenetics – defining the enemy within. *Nat. Med*. **4**, 1107–1111 (1998).

Rafnar, T. et al. The Icelandic Cancer Project – a population-wide approach to studying cancer. *Nat. Rev. Cancer* **4**, 488–492 (2004).

Scadden, D. T. Cancer stem cells refined. *Nat. Immunol*. **5**, 701–703 (2004).

Smallbone, K., Gatenby, R. A., Gillies, R. J., Maini, P. K. & Gavaghan, D. J. Metabolic changes during carcinogenesis: Potential impact on invasiveness. *J. Theor. Biol*. **244**, 703–713 (2007).

Smela, M. E., Currier, S. S., Bailey, E. A. & Essigmann, J. M. The chemistry and biology of aflatoxin B(1): From mutational spectrometry to carcinogenesis. *Carcinogenesis* **22**, 535–545 (2001).

Spencer, S. L. et al. Modeling somatic evolution in tumorigenesis. *PLoS Comput. Biol*. **2**, e108 (2006).

Stein, L. D. Human genome: End of the beginning. *Nature* **431**, 915–916 (2004).

Weinberg, R. A. How cancer arises. *Sci. Am*. **275**, 62–70 (1996).

Willingham, A. T. & Gingeras, T. R. TUF love for 'junk' DNA. *Cell* **125**, 1215–1220 (2006).

Chapter 2
Oncogenes

What is an Oncogene?

An oncogene is a mutated form of a normal cellular gene – called a *proto-oncogene* – that contributes to the development of a cancer. Proto-oncogenes typically regulate cell growth and cell differentiation. Most proto-oncogenes are highly conserved in evolutionarily diverse species, underscoring the fact that genes of this class play central roles in fundamental cellular processes. Mutations of proto-oncogenes that cause their conversion to oncogenes cause many of the perturbations in cell growth and differentiation that are commonly seen in cancer cells.

An oncogene is a type of cancer gene. While all cancer genes are created by mutation, oncogenes are unique in that they are caused by mutations that alter, but do not eliminate, the functions of the proteins they encode. Proteins encoded by oncogenes typically show an increased level of biochemical function as compared with the protein products of the corresponding, nonmutated proto-oncogene.

Most proto-oncogenes encode enzymes. The oncogenic forms of these enzymes have a higher level of activity, either because of an altered affinity for substrate or a loss of regulation. To reflect these gains of function, the mutations that convert proto-oncogenes to oncogenic alleles are known as *activating mutations*.

The Discovery of Transmissible Cancer Genes

The first cancer genes to be discovered were oncogenes. Indeed, the oncogene concept was the first redaction of what would eventually become the cancer gene theory.

Oncogenes were initially discovered as intrinsic components of viruses that cause cancer. Present-day molecular oncologists can trace their scientific lineage to the pioneering virologists of the early 20th century. This group of technologically advanced and elite scientists established many of the laboratory methods and reagents that are essential to modern cancer research. The early virologists created a scientific infrastructure that would facilitate studies of cells and genes. In a tangible

F. Bunz, *Principles of Cancer Genetics*.
© Springer 2008

way, the revolution triggered by the germ theory begat a successive revolution in cancer research.

By the early 20th century, the germ theory was firmly established, as were scientific methods for the systematic study of infectious agents. It was both technically feasible and intellectually compelling to explore whether cancer, like many other common diseases, might have an infectious etiology. Particularly interesting at that time were viruses, which were a new and largely mysterious entity. Viruses were largely uncharacterized, and defined simply as submicroscopic infectious agents present in tissue extracts that would pass through fine filters.

Early experimental observations that laid the foundation for the discovery of oncogenes predated the era of molecular biology. In 1908, Willhelm Ellerman and Olaf Bang demonstrated that a filtered extract devoid of cells and bacteria could transmit leukemia between chickens. Leukemia was not yet recognized as a form of cancer at that time, so this work had little impact. Two years later, Peyton Rous discovered that chicken sarcomas could be serially transmitted from animal to animal by cell-free tumor extracts (Fig. 2.1). The causative agent in the cell filtrates, the Rous sarcoma virus (RSV), was among the first animal viruses to be isolated. The discovery of oncogenic viruses like RSV for the first time led a cancer-causing agent to be studied from a genetic perspective.

The idea that infectious agents cause cancer has a long and tortuous history. The contagious nature of cancer was promulgated in classical times by the widespread belief that cancer was commonly transmitted between individuals in intimate contact with one another, particularly between spouses, from mothers to children and from patients to caregivers. Such beliefs persisted well into the 19th century, when they were gradually disproven by rigorous epidemiology.

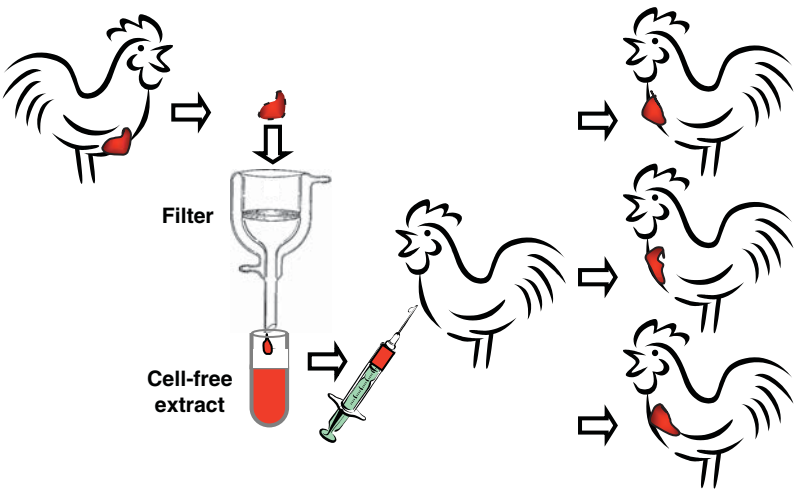

Fig. 2.1 The Rous experiment. A chicken sarcoma extract is prepared by filtration of a homogenized tumor (red). Injection of the cell-free filtrate results in horizontal transfer of the sarcoma to multiple chickens. This experiment demonstrated the infectious nature of this avian cancer

A resurgence of interest in infectious agents as common causes of cancer was prompted by the formulation of the germ theory at the end of the 19th century. Various bacteria, yeasts, fungi, protozoa, spirochetes and coccidia were, at times, briefly implicated as potential agents that could transmit cancer, but subsequent studies failed to support a positive association. As negative results accumulated, the idea that cancer has an infectious etiology fell out of favor once again.

The initial reports by Rous were met with a considerable amount of skepticism. It was suggested that his cell-free filtrates contained active cell fragments or even submicroscopic cells. The prevailing climate of antipathy towards an infectious cause of cancer substantially delayed full acceptance of Rous' work. The idea that viruses could cause cancer was dogmatically rejected as late as the 1950s, despite intermittent reports showing that other cell-free solutions could induce diverse cancers, including breast cancer, in experimental animals. Eventually, the preponderance of evidence grew too large to discount. Peyton Rous was awarded the Nobel Prize in 1966, 55 years after his pioneering work was first published.

Interest in viruses as a cause of human cancer reached a new peak with the discovery of DNA tumor viruses in the 1960s. As the name of this category of viruses suggests, these common papovaviruses could cause tumors in animals and induce cancer-like characteristics in cultured cells. These results led to the resurgence of the idea that viruses might be important to the etiology of human cancer. The contemporary discovery of the DNA tumor virus simian virus 40 (SV40) as a contaminant in polio vaccine stocks that had been previously administered to millions of people was particularly disconcerting. As was the case with other infectious agents that had generated interest in decades past, large follow-up studies failed to establish a causal relationship between the DNA tumor viruses and common human cancers. Despite the fact that they are not a significant cause of cancer, DNA tumor viruses have nonetheless been very useful tools for cancer research. The most widely mutated gene in human cancer, *P53*, was initially discovered by virtue of its physical association with an SV40 viral protein in cultured cells (see Chapter 3).

As discussed in Chapter 1, the viruses that have a measurable impact on the incidence of human cancer typically stimulate a chronic inflammatory response. Inflammation, in turn, creates a microenvironment that promotes the acquisition, by mutation, of cancer genes and the proliferation of cells that harbor cancer genes. The DNA tumor viruses do not fall into this category and are not considered carcinogenic.

There is no known virus that causes cancer in humans in the dramatic way that RSV causes cancers in chickens. Nonetheless, the use of RSV to induce chicken tumors provided an invaluable model system that showed how a simple genetic element could cause cells to acquire cancer phenotypes. Prior to the completion of the human genomic sequencing draft released in 2000, most of the human genome was for practical purposes a black box. The information contained in the genome as a whole was largely unavailable or inaccessible. Cancer-associated viruses presented researchers with relatively short, well-defined regions of DNA sequence that were known to directly relate to cancer development. Viral genes could be fully sequenced and experimentally manipulated with recombinant DNA technology that

was developed in the 1970s and the 1980s. The unraveling of the complex relationship between the genes of cancer-associated viruses and human genes was a pivotal step in the elucidation of the cancer gene theory.

Viral Oncogenes are Derived from the Host Genome

The sarcoma virus isolated by Rous is one of the most potent carcinogens known. Inoculation of chickens with RSV results in the appearance of tumors within several weeks. This acute onset is in stark contrast to the development of most human tumors, which take decades to develop. Clearly, viruses like RSV have evolved a unique mechanism to trigger the cellular changes that cause cancer.

RSV belongs to a category of viruses now known as the *retroviruses*. Retrovirus particles contain genomes that are in the form of ribonucleic acid (RNA). After infection with RSV, the retroviral RNA genome is copied into DNA by the virus-encoded enzyme *reverse transcriptase*. The viral DNA then integrates into the host genome, and thus becomes what is known as a *provirus*. The provirus is replicated along with the host genome by the host DNA replication machinery, and is also transcribed by host RNA polymerase complexes. The proviral RNA transcripts are packaged into new virions, completing the virus life cycle (see Fig. 2.2).

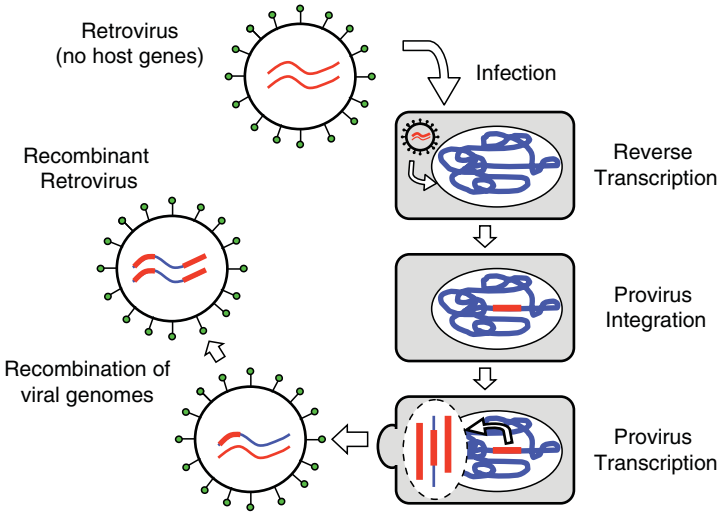

Fig. 2.2 The acquisition of oncogenes by retroviruses. The retrovirus capsule contains two copies of the viral RNA genome. After infection, the viral genome is copied into DNA by reverse transcriptase and integrates into the cellular genome as a provirus. If the provirus is integrated in close proximity to exon sequences, proviral transcripts can be spliced with host cell exons. These hybrid transcripts are packaged into a virion, resulting in a heterozygous viral genome. The viral genome undergoes recombination during a second round of infection. The resulting recombinant virus contains coding genetic elements that originated in the host cell

Retroviruses can cause cancer in two different ways. Depending upon where they integrate, proviruses can disrupt the functions of host genes, usually by altering their transcriptional regulation. In effect, a proto-oncogene can be changed into an oncogene upon integration of a provirus. Typically, cancers caused by the disruption of a host gene by a provirus have a long latent period and take a long time to develop. The viruses that cause such tumors are accordingly known as *slowly transforming retroviruses*. In contrast, acutely transforming retroviruses such as RSV carry their own cancer genes.

RSV contains a cancer gene known as *SRC* (pronounced 'sark'). The protein encoded by *SRC* is an enzyme that localizes near the cell membrane and covalently modifies proteins in response to growth signals (see Fig. 2.3). Specifically, *SRC* encodes a protein tyrosine kinase, a class of enzymes that catalyzes the addition of a phosphate group onto the tyrosine residues of multiple protein substrates, thereby altering their function. Each covalent modification catalyzed by the *SRC*-encoded protein is one event of a series of enzymatically controlled events that collectively function to mediate signals that promote cell growth and division. In short, the *SRC*-encoded protein signals the cell to grow. The biochemical modes by which the enzymes encoded by cancer genes act as cellular messengers will be discussed in detail in Chapter 5.

In a landmark paper published in 1976, J. Michael Bishop, Harold Varmus and their colleagues demonstrated that the retroviral genes that cause avian cancers are actually variants of genes present in the chicken genome. There are in effect

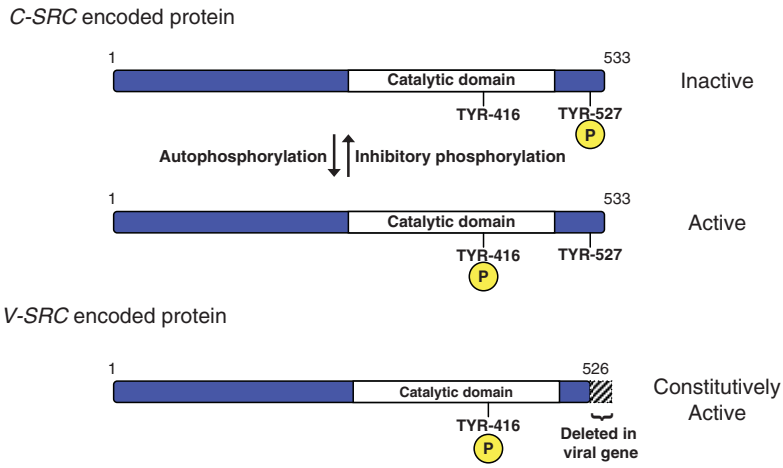

Fig. 2.3 Viral and cellular *SRC* genes. Cellular *SRC* (*C-SRC*) is a protein tyrosine kinase, 533 amino acid in length. Tyrosine autophosphorylation at residue 416 within the kinase domain causes a conformational change in the protein that results in the activation of kinase activity. Phosphorylation at tyrosine 527 by upstream inhibitory kinases prevents *C-SRC*-encoded protein activation. The viral oncogene *V-SRC* does not encode the c-terminal seven amino acids, and therefore does not contain the negative regulatory element

two related *SRC* genes. The cellular form of the *SRC* gene, denoted *C-SRC*, is a proto-oncogene that encodes a protein containing a tyrosine residue in the carboxy-terminus. This residue is a substrate of an enzyme that regulates growth in concert with the C-SRC protein (see Chapter 5). When phosphorylated at this tyrosine residue, the *C-SRC*-encoded protein is rendered functionally inactive and does not transduce growth signals. In contrast, the *SRC* gene carried by RSV, *V-SRC*, encodes a protein that has a truncated carboxy-terminus, and therefore does not contain the tyrosine residue that is the target of the inhibitory signal. The *V-SRC*-encoded protein thus is missing a regulatory feature present in the *C-SRC*- encoded protein. The role of C-SRC and protein phosphorylation in cancer is described in detail in Chapter 5.

How did a host gene come to reside in a retrovirus? The answer lies in the retrovirus life cycle (see Fig. 2.2), during which retroviruses shuttle in and out of the host genome. It appears that retroviruses acquired cellular genetic material over the course of these cycles by recombination of the viral DNA with cellular DNA, and incorporated these genes into their own genomes. Evolutionary forces would favor proviruses that can most effectively propagate. Once integrated, the fate of a provirus becomes linked to the fate of the host cell genome. Proviruses that contain genes such as *V-SRC* trigger DNA replication and cell proliferation and thereby promote their own production.

The observation that cancer-causing retroviruses contain altered forms of host genes was a watershed event that fundamentally changed the focus of cancer research. This critical finding showed that the key to understanding cancer lies in the genome of the cancer cell itself. For the first time it was clear that altered cellular genes could cause cancer.

The Search for Activated Oncogenes: The RAS Gene Family

The oncogenes that most often contribute to the development of human cancers are not transmitted by viruses, but rather are acquired by the somatic mutation of proto-oncogenes. The horizontal transfer of cancer by RSV-containing cell extracts does not reflect the means by which human cells acquire oncogenes. Nonetheless, viruses such as RSV did provide important insight as to what oncogenes look like and to how they might induce cellular changes.

The idea that oncogenes could be transmitted by some viruses fostered creative strategies to isolate additional genes that might have oncogenic potential. Genetic material can be efficiently transferred to cultured cells by chemical techniques that were developed during the 1970s. When introduced into primary cells growing in culture dishes, oncogenes can cause observable changes in growth properties. In a process known as *in vitro transformation*, cells that are experimentally forced to express many types of oncogene undergo changes in morphology, lose contact inhibition and begin to grow in piles known as *foci* (see Fig. 2.4). These quantifiable changes formed the basis of numerous experiments that led to the discovery of several widely mutated oncogenes.

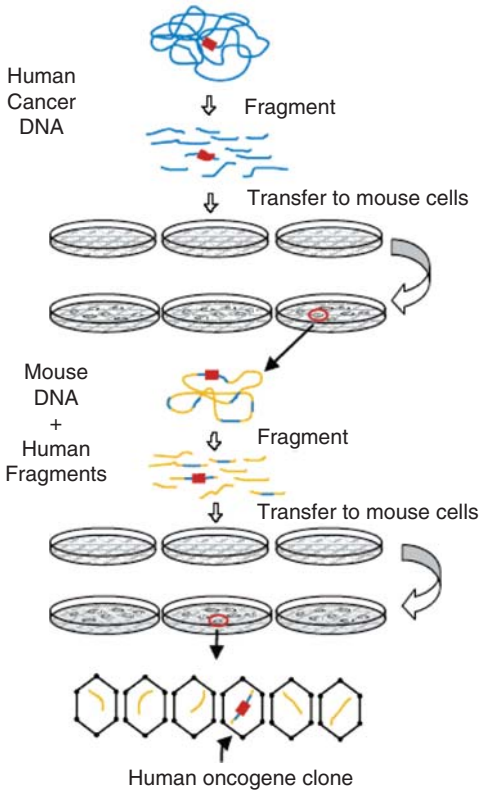

Fig. 2.4 Oncogene discovery by *in vitro* transformation. Genes transferred from human genomic DNA (blue) can alter the growth properties of mouse fibroblasts. Genomic DNA is sheared into smaller fragments, which are introduced into mouse cells growing in monolayer cultures. Appearing after a period of growth, discrete foci represent clones of mouse cells that have altered growth and cell–cell interactions. Genomic DNA from these clones (yellow) can contain multiple integrated fragments of human DNA. A second round of transfer allows the isolation of individual human fragments. DNA from the second clone is packaged into a bacteriophage library, which is then screened with a probe corresponding to human genomic DNA-specific repeat elements. Assays of this type were relatively nonspecific. Foci can be caused by actual oncogenes that are activated in cancer cells, but also by proto-oncogenes activated by the gene transfer process and growth regulatory genes that are not found to be mutated in cancers

Potent oncogenes were found to be carried by two retroviral strains, the murine Harvey and Kirsten sarcoma viruses. These *r*etrovirus-*a*ssociated DNA sequences (or RAS genes) were designated *H-RAS* and *K-RAS*, respectively. The Harvey and Kirsten retroviruses were not naturally occurring pathogens, but had been experimentally derived by repeated passage of murine leukemia viruses through laboratory strains of rats. During the creation of these new, highly carcinogenic viruses, *H-RAS* and *K-RAS* had been acquired in altered, oncogenic form from the host genome. Using DNA transfer schemes, the laboratories of Robert Weinberg, Geoffrey

Cooper, and of Mariano Barbacid and Stuart Aaronson independently isolated variants of the RAS gene family directly from human cancer cells.

That retroviral oncogenes are related to the oncogenes created by the somatic mutation of proto-oncogenes was underscored by the discovery of the RAS genes. Activated RAS alleles were the first cancer genes to be found in cells derived from naturally occurring human cancers. It was shown that the RAS genes isolated from human bladder and lung carcinoma cells were homologous to the RAS genes harbored by the Harvey and Kirsten retroviruses. Soon thereafter, Michael Wigler and colleagues isolated a third RAS gene family member, that had no known viral homolog, from a neuroblastoma. The third RAS gene was designated *N-RAS*. These three genes are encoded by distinct loci but are highly related, both structurally and functionally.

The wild type RAS proto-oncogenes do not induce focus formation in the in vitro transformation assay. The gain of function that leads to the acquisition of this property is conferred by an activating point mutation. For example, the bladder carcinoma from which the cellular *H-RAS* gene was first isolated was found to have a single base substitution that changed codon 12 from GGC (glycine) $\rightarrow$ GTC (valine). Subsequent DNA sequence analysis of large numbers of human tumors has revealed a high frequency of RAS gene mutations in several tumor types. The majority of these cancer-associated mutations involve just three codons: 12, 13 and 61. Different tumor types differ greatly in the overall frequency of RAS gene mutations, and also in the RAS family member that is predominantly mutated (see Table 2.1).

Interestingly, the first *RAS* oncogenes discovered were not representative of naturally occurring activated oncogenes. Although activated *H-RAS* was among the first oncogenes to be discovered in a tumor, mutations in this RAS family member are not widespread in cancers. Similarly, *N-RAS* was first isolated from a neuroblastoma, yet subsequent studies have failed to detect *N-RAS* mutations in a significant proportion of these tumors. It remains a possibility that the mutated RAS genes identified by in vitro transformation arose during the maintenance of tumor-derived cell lines in culture (in vitro), rather than by somatic mutation that occurred during tumorigenesis. Nonetheless, the initial identification of the RAS family of oncogenes was an important achievement that paved the way for the systematic analysis of common cancer mutations. Mutations in *RAS* family members are involved in a significant proportion of a number of common malignancies (see Table 2.1).

RAS genes are ubiquitously expressed and presumably have the same function in all cells. Why then is mutation of *K-RAS* a dominant feature of pancreatic tumors

Table 2.1 Mutations in the *RAS* gene family

Cancer type	Mutation frequency (%)	RAS family member
Pancreatic carcinoma	95	*K-RAS*
Colorectal carcinoma	50	*K-RAS*
Lung carcinoma	30	*K-RAS*
Acute Myelogenous Leukemia	25	*N-RAS*
Melanoma	10	*N-RAS*

and present at much lower frequencies other malignancies? Why are *N-RAS* mutations but not other RAS family mutations prevalent in acute myelogenous leukemias? The basis for the tissue specificity of RAS mutations, and indeed of cancer gene mutations in general, is largely unknown. One might assume that tissue-specific gene alterations arise in cancers at a detectable frequency because they provide a selective advantage in a given cellular compartment.

The cellular role of the RAS-encoded proteins involves the coupling of signals that arise at cell membrane receptors with downstream intracellular signaling molecules. RAS proteins are therefore frequently described as second messengers. The mutation of conserved codons in the RAS family members affects the regulation of the enzymatic activity of RAS proteins. The nature of RAS protein activity and the cellular functions of the RAS gene family will be discussed in detail in Chapter 5.

Complex Genomic Rearrangements: The MYC Gene Family

The MYC gene family first emerged as a viral gene, *V-MYC*, harbored in the genomes of four independent isolates of avian leukemia virus. Among the tumors caused by these oncogenic retroviruses is myelocytomatosis, a tumor composed mainly of myelocytes, a type of white blood cell. It is from this rare tumor that the name of a commonly activated oncogene family was derived. The cellular homolog of *V-MYC* is the proto-oncogene *C-MYC*. There exist two structurally and functionally related genes that were discovered subsequently, designated *N-MYC* and *L-MYC*. The latter two genes were isolated as oncogenes from a neuroblastoma and a lung carcinoma, respectively.

In contrast to the genes in the *RAS* family, which are activated by single nucleotide substitutions, MYC genes are typically activated by larger and more complex genomic rearrangements. The encoded protein product is not structurally altered by MYC gene activation, but increased in quantity. The consequence of MYC activation, regardless of the precise mechanism, is an increase in gene expression. Even modest increases in MYC gene expression caused by activating mutations are thought to significantly contribute to tumorigenesis in some tissues.

The MYC genes encode transcription factors that directly affect the expression of genes involved in several aspects of cell growth as it relates to tumor development and progression. The MYC genes are sometimes referred to as nuclear proto-oncogenes, reflecting their role in controlling the transcription of genes in the cell nucleus. The function of the MYC genes in the alteration of gene expression in cancer cells will be discussed in Chapter 5.

The three MYC genes share a common genomic structure that consists of 3 exons. Including intronic regions, each spans approximately 5 kb. This compact genetic unit has been found to be rearranged in a number of ways that result in the aberrantly high expression of MYC proteins. Studies of MYC genes in cancers have revealed several general mechanisms by which proto-oncogenes can be activated.

All of the activating mutations that convert MYC genes to their oncogenic forms increase the protein levels. There are several mechanisms by which this occurs. The number of functional MYC genes can increase as a result of the amplification of the genomic region containing a MYC gene. Alternatively, the level at which a MYC gene is expressed can be altered if that gene is repositioned in proximity to a highly active promoter element, usually as a result of a chromosomal translocation. These genetic changes are types of somatic mutations that are stably propagated by cancer cell clones during their evolution.

Proto-Oncogene Activation by Gene Amplification

In normal cells, proto-oncogenes exist as single copy genes. That is, a single genomic locus contains one copy of each exon, intron and regulatory element. Due to the diploid nature of the human genome, a total of two copies of each gene will be present in each cell, one on each of the two homologous chromosomes.

The copy number of a gene can increase as a result of the amplification of a subchromosomal region of DNA. The increase in gene copy number leads, in turn, to a corresponding increase in the overall expression levels of that gene. The process by which genomic amplification occurs remains incompletely understood, but is thought to involve repeated rounds of DNA replication that occur during a single cell cycle.

The unit of genomic DNA that is amplified is known as the *amplicon*. Amplicons vary in size, but typically range in size between 10^5 and 10^6 base pairs. The number of amplicons found within a region of amplification also varies broadly. An amplicon can contain varying numbers of genes depending on the size and location of the genomic region contained within the amplicon. Overall genomic structure is typically preserved within amplified regions, with amplicons ordered in repetitive arrays in head-to-tail orientation (see Fig. 2.5).

If the copy number is high or if an amplicon is particularly large, the amplified region may be microscopic and therefore directly observable by cytogenetic methods (see Fig. 2.6). Amplified regions of the genome can exist in extrachromosomal bodies known as *double minutes*, which are small structures that resemble chromosomes but do not contain centromeres. Double minutes can integrate into a chromosome. The region of integration can often be distinguished cytogenetically as a region that stains homogenously with dyes used to reveal chromosome banding patterns. The integration of double minutes is thought to be reversible. Accordingly, the integrated and extrachromosomal forms of amplified genomic DNA are believed to be interchangeable. Double minutes and homogeneous staining regions are not seen in normal cells upon cytogenetic analysis, but are seen in a significant number of tumor cells.

Upon amplification of a MYC locus, MYC is converted from a proto-oncogene to an oncogene. The most notable role for *N-MYC* amplification is in the growth of neuroblastomas, tumors that arise from immature nerve cells. These tumors almost

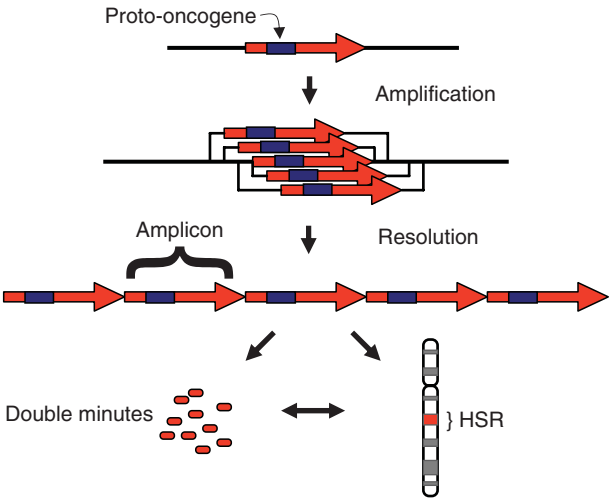

Fig. 2.5 Oncogene activation by gene amplification. A genomic region (red arrow) containing a proto-oncogene is amplified as a result of multiple rounds of DNA replication during a single cell cycle. Resolution of the over-replicated region results in a tandem array of amplicons in head-to-tail orientation. The amplified region can alternatively be maintained as double minutes, or integrated into a chromosome to form a heterogenous staining region (HSR). It is believed that these two configurations are interchangeable

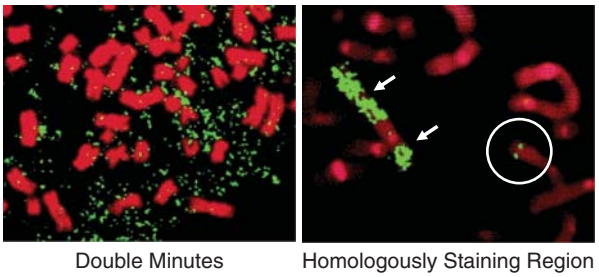

Fig. 2.6 Amplified C-MYC. The *MYC* locus in mitotic cells is stained green by fluorescence in situ hybridization. Shown as left are double minutes containing the amplified *C-MYC* locus. In the right panel are two homologously staining regions, indicated by arrows. Circled in the same panel are the two endogenous, unamplified *C-MYC* loci. (From Savalyeva and Schwab, *Cancer Lett.* **167**, 115–123 (2001). With permission.)

exclusively affect young children. Amplification of the genomic region on chromosome 2p24 containing *N-MYC* can be detected in about 25% of neuroblastomas. The degree of amplification of *N-MYC* in neuroblastomas can be extensive; as many as 250 copies have been found in some of these cancers. The extent of *N-MYC* amplification has been found to correlate with both the stage of the disease, and independently with the rate of disease progression and outcome. These findings

provide evidence that *N-MYC* amplification directly contributes to neuroblastoma progression.

Amplified MYC genes are commonly found in a number of tumors in addition to neuroblastomas. The first example of *C-MYC* amplification was observed in a myelocytic leukemia. *C-MYC* amplification is frequently observed in cervical cancers and esophageal cancers. Small-cell cancers of the lung have been found to variously contain amplification of one of the three MYC genes, *C-MYC, N-MYC* and *L-MYC*. *C-MYC* amplification is found in approximately 20–30% of breast carcinomas and appears to be correlated with a poor clinical outcome.

Another gene that is commonly amplified in a broad spectrum of cancers is *ERBB2*, alternatively referred to as *HER2/neu. ERBB2* amplification has been found in a significant proportion of breast and ovarian cancers and also in adenocarcinomas arising in the stomach, kidneys and salivary glands.

The *ERBB2* gene was first identified as the cellular homolog of an oncogene, *VERBB2*, carried by the avian erythroblastic leukemia virus, a retrovirus. At around the same time, an oncogene termed *NEU* was isolated from a rat neuroblastoma cell line by in vitro transformation, while a gene known as *HER2* was discovered by virtue of its similarity to a previously discovered gene that encodes a cell surface signaling protein called human epidermal growth factor receptor. Efforts to determine the chromosomal locations of these genes suggested – and DNA sequencing subsequently proved – that *HER2/neu* and *ERBB2* are in fact a single gene.

Genetic alterations that activate *ERBB2* are among the most common somatic mutations found in breast cancer, occurring in approximately 15–25% of tumors analyzed. The majority of these are gene amplifications that result in increased *ERBB2* expression. The amplicons that include the entire *ERBB2* locus vary between cancers but span a common region of about 280 kb in length. This core amplicon includes several loci in addition to *ERBB2*, but genetic analysis strongly suggests that it is the enhanced expression of *ERBB2* that confers clonal selectivity. Amplified regions typically contain about 20 copies of the *ERBB2* amplicon, but have been found to contain as many as 500 copies. Analysis of the *ERBB2* coding regions has not revealed any alterations that affect the open reading frame, confirming that the increase in gene dosage is the most probable activating factor.

ERBB2 encodes a protein that functions as a receptor on the cell surface that transduces growth signals. The activation, by amplification, of this proto-oncogene results in the overexpression of the ERBB2 receptor and a resulting hypersensitivity to growth factors. The *ERRB2*-encoded protein is a prototype of an important class of oncogene-encoded proteins that will be described further in Chapter 5.

Amplification of *ERBB2* in breast cancers is a useful prognostic marker. While amplification of *ERBB2* does not appear to correlate with disease characteristics such as tumor size, there is a significant correlation with the spread of cancer cells to local lymph nodes, which is independently a negative prognostic sign. Breast tumors that harbor *ERBB2* amplification tend to grow more aggressively. Statistically, patients with *ERBB2* positive cancers exhibit a significantly shorter time to relapse following standard therapy and reduced long-term survival. The recent development of specific therapy that targets *ERBB2* function makes the identification of

Table 2.2 Oncogenes frequently amplified in human cancers

Oncogene	Cellular function	Type of cancer	%
C-MYC	Transcription factor	Cervical	25–40
		Esophageal	38
		Breast	20
		Non-small cell lung	15
CCND1	Cell cycle regulator	Head and neck	50
		Breast	20
		Esophageal	25
		Hepatocellular	13
CCNE	Cell cycle regulator	Gastric	15
CDK4	Cell cycle regulator	Sarcoma	11–80*
		Glioblastoma	15
EGFR (ERBB1)	Growth factor receptor	Glioblastoma	33–50
ERBB2	Growth factor receptor	Medulloblastoma	40
(HER2/neu)		Breast	20–35
		Ovarian	20
		Cervical	20
		Non-small cell lung	10
HDM2	Regulation of tumor suppressor protein	Sarcoma	10–90*
MET	Protein tyrosine kinase	Esophageal	80
		Medulloblastoma	40
		Gastric	10–20
MITF	Transcription factor	Melanoma	20
PIK3CA	Lipid kinase	Medulloblastoma	45
		Ovarian	15

*Varies depending on cell type of origin.

patients with *ERBB2* overexpressing tumors a priority. The molecular basis for targeted therapies is discussed in Chapter 7.

Oncogenes activated by gene amplification contribute to many common types of cancer (see Table 2.2).

Proto-Oncogene Activation by Chromosomal Translocation

A chromosomal break presents a unique challenge to a growing cell. Cells that contain broken chromosomes cannot continue to grow and divide; proliferation can only continue once a chromosomal break is repaired. As described in Chapter 1, there are several mechanisms that can function to mend a double strand DNA break and thus repair a broken chromosome. The resolution of such breaks is critical to cell survival, but the process of repair frequently results in mutations. One such mutation is the *chromosomal translocation*.

A translocation is the transfer of a chromosome segment to a different position, often on a nonhomologous chromosome. In some cases the repair process results in the exchange of pieces between nonhomologous chromosomes; such an exchange is termed a reciprocal translocation (see Chapter 1).

Gross structural rearrangements like translocations can juxtapose proto-oncogenes with genetic elements that normally would be distant. Proto-oncogenes can be activated by translocations in two ways, depending on the location of the break point. A translocation can put the exons of two separate genes under the control of a single promoter element. This intermingling of exons can then result in the expression of a single fusion protein that contains elements of each of the two genes involved. Alternatively, a translocation can preserve a complete open reading frame but juxtapose the coding exons with a highly active promoter.

An example of a proto-oncogene that can be activated by chromosomal translocation is *C-MYC*. The expression of *C-MYC* is normally tightly regulated. This tight transcriptional control is altered in some lymphomas and leukemia in which the *C-MYC* gene is repositioned, via translocation, into the vicinity of a highly active promoter. The repositioning of *C-MYC* into the vicinity of these strong promoters is sufficient to activate *C-MYC*, and thereby convert it into a functional oncogene.

Chromosomal Translocations in Liquid and Solid Tumors

Somatically acquired chromosomal translocations are frequently found in the liquid tumors: the leukemias and lymphomas. Although translocated chromosomes have been found in many solid tumors, the more common translocations found in the liquid tumors are tightly associated with specific disease. Translocations that convert proto-oncogenes to oncogenes have been found in over 50% of leukemias and in a significant proportion of lymphomas.

Some common genetic alterations are repeatedly observed in cancers of a single type from many different patients. Such alterations are said to be *recurrent*. Many of the recurrent translocations found in liquid tumors are structurally conserved and defined by common break points. These break points often occur in closely spaced clusters. The location of the break points or break-point clusters that define translocations is highly disease-specific and in some cases diagnostic.

Cancers that arise in particular cell type will typically harbor similar translocations. As specific types of mutations are associated with subsets of solid tumors, the oncogenes located near break points are specifically activated in certain subsets of liquid tumors.

Recurrent translocations, like other genetic alterations, are lineage dependent. The recurrent translocations involving *C-MYC* indicate why this is the case. The chromosomal translocation resulting in the juxtaposition of *C-MYC* and highly expressed immunoglobulin genes is a common feature of both B-cell leukemia and Burkitt lymphomas, particularly those arising in children. These cancers arise from a common stem cell, the lymphoid progenitor, in which immunoglobulin

gene expression is highly activated. In contrast, *C-MYC* is activated in T-cell leukemias by translocation and juxtaposition with highly expressed T-cell receptor genes. In these distinct cancers, both the oncogene and the mode by which is it activated are recurrent. Additionally one would readily infer that *C-MYC* activation confers a particularly strong survival advantage in these distinct tissue compartments.

Despite the fact that solid tumors are much more common than liquid tumors, less is known about the overall role of chromosomal translocation in solid tumors. This paucity of information may be partly due to the technical obstacles that are inherent to analyzing chromosomes in solid tissues. Cytogenetic analysis is considerably more difficult in solid tumor samples for several reasons. Solid tumors grow and develop over a considerable length of time. Often, decades elapse during the evolution of a large, invasive tumor from a small neoplasm. During this time, tumors can become heterogeneous. Portions of tumors that are starved of oxygen and nutrients can die by the process of necrosis. Dead or dying cancer cells, infiltrating inflammatory cells and cells from adjacent normal tissues are present in varying proportions in biopsy samples and can complicate cytological analysis. As a result of these complications, recurrent chromosomal translocations and their contribution to cancer development remain best understood in leukemias and lymphomas.

Chronic Myeloid Leukemia and the Philadelphia Chromosome

The activation of a proto-oncogene by a pathognomonic translocation is best illustrated by the example of chronic myeloid leukemia (CML). In 95% of CML patients, the cancer cells contain a unique derivative chromosome named after the city in which it was discovered, the Philadelphia chromosome (see Fig. 2.7). The Philadelphia chromosome was originally identified in 1960 and upon detailed cytogenetic analysis in 1973 was found to result from a reciprocal translocation involving chromosomes 9 and 22. Five percent of CML patients that do not exhibit a typical Philadelphia chromosome have translocations that are structurally more complex, but still ultimately involve the same chromosomal regions. Subsequent to its discovery in CML patients, the Philadelphia chromosome was also found to be present in 3–5% of children and 30–40% of adults with acute lympohcytic leukemia (ALL).

CML is a cancer that arises in blood cell progenitors and spreads throughout peripheral blood and bone marrow. CML affects all age groups, but is most common in older adults. The natural history of CML unfolds in clinically defined stages. Within 3–5 years after its detection, CML typically progresses from a relatively benign chronic disease to an acute illness – known as blast crisis – that is frequently fatal. While the CML cells found during the chronic stage are mature, those found during blast crisis are relatively undifferentiated and resemble those found in patients with acute leukemias.

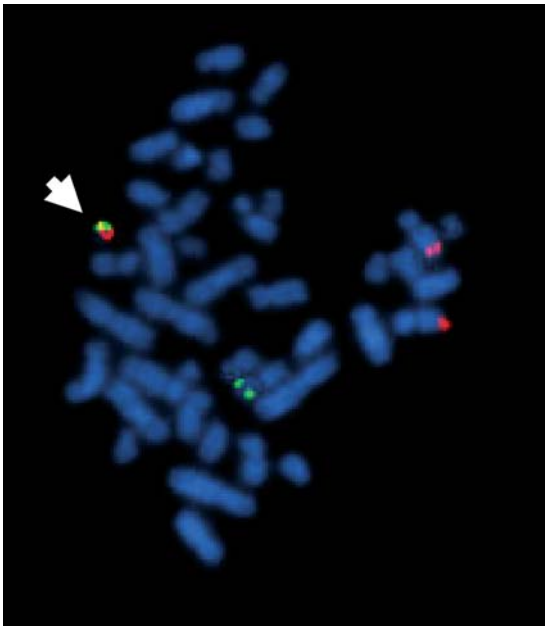

Fig. 2.7 The Philadelphia chromosome. The Philadelphia chromosome (indicated by arrow) stained during mitosis. Fluorescence in situ hybridization probes are derived from *BCR* (green) and *C-ABL* (red). The spots in other chromosomes represent the untranslocated *BCR* and *C-ABL* genes

Interestingly, the only environmental factor known to predispose people to CML is exposure to ionizing radiation. It is possible that the repair of double strand DNA breaks caused by ionizing radiation results in the stochastic generation of the Philadelphia chromosome. In most cases, no predisposing factors are identified and as a result the cause of the initial translocation is usually obscure. Regardless of the mechanism by which they arise, the rare cells containing this translocation are then clonally selected and expanded by the process of clonal evolution (Chapter 1). The recurrence of a single translocation in CML suggests that this genetic alteration must provide the cancer precursor cells with a unique and essential survival advantage.

At the molecular level, the consequence of the translocation involving chromosomes 9 and 22, denoted t(9;22), is the unique juxtaposition of two genes, *BCR* and *C-ABL*. *C-ABL* is a proto-oncogene homologous to an oncogene originally found in the retroviral genome of the Ableson leukemia virus. In the absence of translocation, the expression of the *C-ABL* proto-oncogene is tightly regulated. The *BCR* gene, in contrast, was so named because of its location within the *b*reak point *c*luster *r*egion on chromosome 22. *BCR* expression is driven by a strong, constitutively active promoter. Strictly speaking, *BCR* is not considered a proto-oncogene, and in fact its normal cellular role is unknown. The *BCR* promoter functions to transcribe *C-ABL* exons when the two genes are fused by translocation (see Fig. 2.8).

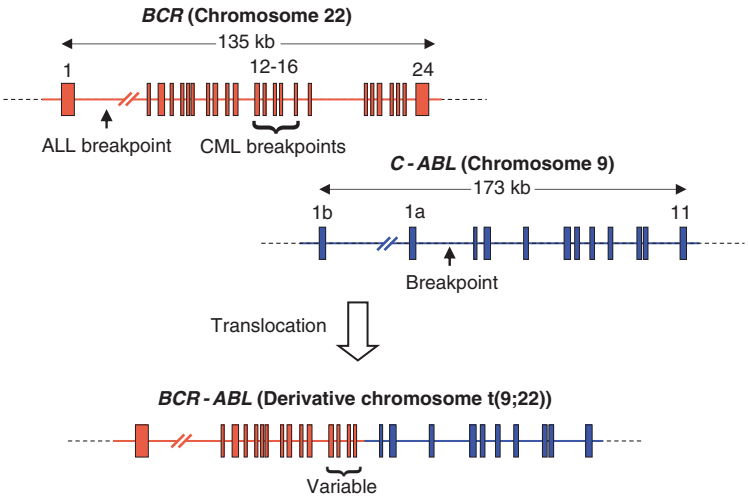

Fig. 2.8 The creation of *BCR-ABL* by translocation. The BCR locus on chromosome 22 spans roughly 135 kb and is composed of 24 exons. Within this gene is a recurring break point found in acute lymphocytic leukemia-associated translocations, and a cluster of break points found in chronic myeloid leukemias. The C-ABL locus on chromosome 9 spans 173 kb and has 11 exons. Note that there are two first exons that are alternatively utilized. A single recurrent break point occurs upstream of exon 2. In the t(9;22) derivative, the *BCR* and *C-ABL* genes are fused, and contain a single open reading frame. The different CML-associated break points in *BCR* result in the variable inclusion of *BCR* exons 12–15 in different allelic forms of *BCR-ABL*

The t(9;22) reciprocal translocation results in the creation of two separate fusions between the *BCR* and *C-ABL* genes. The *BCR-ABL* gene is created on the derivative of chromosome 22, the Philadelphia chromosome, while a corresponding *ABL-BCR* fusion gene is created on the derivative chromosome 9. Numerous experiments have demonstrated that it is the product of the *BCR-ABL* gene that is oncogenic. Like a substantial number of proto-oncogenes, the *C-ABL* gene encodes a protein tyrosine kinase. The fusion gene encodes the catalytic domain of this enzyme, while the expression of this domain is controlled by the *BCR* promoter. It appears that the BCR peptide mediates oligomerization of the BCR-ABL fusion protein, causing constitutive activation of the protein tyrosine kinase domain in the ABL peptide. The mutational activation of tyrosine kinases and their roles in the cell are discussed in detail in Chapter 5.

The precise junction between chromosome 9 and chromosome 22 sequences varies between different groups of CML patients. While there is a single break point on chromosome 9, the break point on chromosome 22 is actually of cluster of distinct break points variably found in different groups of patients. Accordingly, the portion of *BCR-ABL* that is composed of *C-ABL* sequence is invariant. However, the existence of multiple break points within the *BCR* locus results in the creation

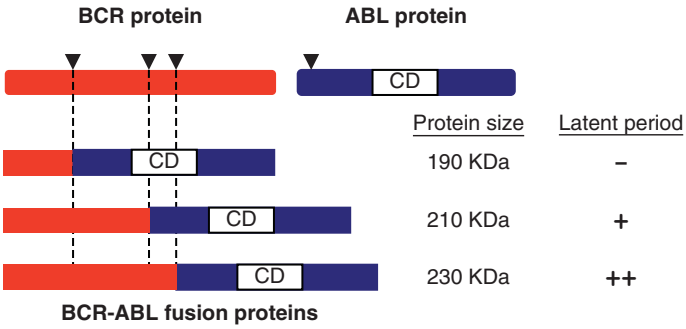

Fig. 2.9 *BCR-ABL*-encoded proteins. The primary structures of the native BCR and ABL proteins are shown. Arrowheads indicate the regions of defined by the recurrent break points. The various break points in *BCR* lead to the appearance of distinct fusion proteins with molecular weights of 190, 210 and 230 kDa. The 190 KDa protein is restricted to ALL, an acute disease that is not characterized by a latent period. The 210 KDa is the most prevalent CML-associated version, while the 230 KDa protein is found in a subset of CML patients that typically exhibit an extended period of disease latency

of distinct in-frame fusions. The chimeric proteins encoded by these different gene fusions differ at their N-termini and can be distinguished by their molecular weight (see Fig. 2.9). Since CML, like all cancers, is monoclonal in nature, only one *BCR-ABL*-encoded protein is detectable in each patient. Depending on the site of the break point in the *BCR* gene, the fusion protein can vary in size from 185 to 230 kDa.

The different *BCR-ABL* fusion proteins can be correlated with different clinical outcomes. Most CML patients express the 210 KDa form of the fusion protein. A subgroup of CML patients has been identified that express a 230 KDa BCR-ABL-encoded protein. These patients have a distinct disease course that is typified by decreased numbers of white cells in the peripheral blood and delayed progression to blast crisis. Patients with highly aggressive ALL express either the 210 KDa form or a unique 190 KDa protein. The 190 KDa protein has been shown to be a more active tyrosine kinase than the 210 KDa protein, suggesting that different levels of activity affect the clinical course of these diseases.

Because the presence of the various gene fusion products correlates with both the type and course of disease, these molecules are useful markers for diagnosis and prognosis. The presence of a chimeric RNA species transcribed from a fusion gene is readily detectable by commonly employed RNA/DNA amplification techniques. Thus, the expression of these unique oncogenes provides a convenient and highly informative marker than can be directly used in the clinic.

The catalytic activity of the *BCR-ABL*-encoded tyrosine kinases can be directly inhibited by drugs. Therapy based on this approach has been highly successful at delaying blast crisis and has significantly improved the overall outlook for patients with CML. The fact that specific therapy directed at the *BCR-ABL* gene product is highly effective demonstrates conclusively the central role of the *BCR-ABL*

oncogene in CML pathogenesis. The affects of tyrosine kinase activation on cancer cell proliferation will be discussed in Chapter 5; novel therapeutic approaches to specifically target these enzymes will be described in Chapter 7.

Ewing's Sarcoma and the Oncogenic Activation of a Transcription Factor

Recurrent chromosomal translocations characterized at the molecular level have not been described in the most common epithelial malignancies, but are found in less common solid tumors. In sarcomas (cancers that arise in connective tissues), specific genetic alterations have been found to be associated with tumor-specific translocations. The role of a chromosomal translocation in the pathogenesis of a solid tumor is illustrated by the example of Ewing's sarcoma.

Ewing's sarcoma is a rare tumor that occurs in children and young adults, most commonly in male teenagers. These highly aggressive tumors can occur in various anatomic sites, but most frequently are seen in bone. The cells that compose Ewing's sarcomas are morphologically similar to those found in diverse types of pediatric solid tumors, making accurate diagnosis difficult. This challenge prompted focused investigation into cytogenetic changes that could potentially provide a diagnostically useful marker. A distinguishing characteristic of Ewing's sarcoma cells was found to be the presence of a reciprocal translocation between chromosomes 11 and 22, abbreviated t(11;22).

Molecular analysis revealed that t(11;22) consistently juxtaposes the *FLI1* gene on chromosome 11 and the *EWS* gene on chromosome 22 (see Fig. 2.10). *FLI1* was originally identified in mice as the integration site common to two retroviruses that cause leukemias and sarcomas, including the Friend leukemia virus for which the locus was named. Human *FLI1* is highly similar to a proto-oncogene called *ETS1*, the cellular homolog of a retroviral oncogene carried by the avian leucosis virus. The encoded proteins of both *FLI1* and *ETS1* share a protein domain that is important for sequence-specific DNA binding, and both proteins are now recognized to belong to a family of related transcription factors. The roles of oncogenic transcription factors in the cancer cell are described in Chapter 5. The *EWS* protein product contains a putative RNA-binding domain, but the normal function of this protein is unknown.

In the Ewing's sarcoma translocation, the chromosomal break points occur within the introns of *FLI1* and *EWS*, and result in the in-frame fusion of the promoter and upstream elements of *EWS* and the downstream elements of *FLI1*. The precise locations of the break points vary from tumor to tumor. The most frequent junction, occurring in 60% of cases, joins exon 7 of EWS to exon 6 of *FLI1* in what is termed a Type 1 fusion. Approximately 25% of cases are associated with a so-called Type 2 fusion, which includes exon 5 of *FLI1*. As was seen to be the case in CML, the fusion variants correlate with distinct clinical outcomes. In particular, the Type 1 fusion is associated with a significantly better prognosis, and specifically better survival, than the other fusion types.

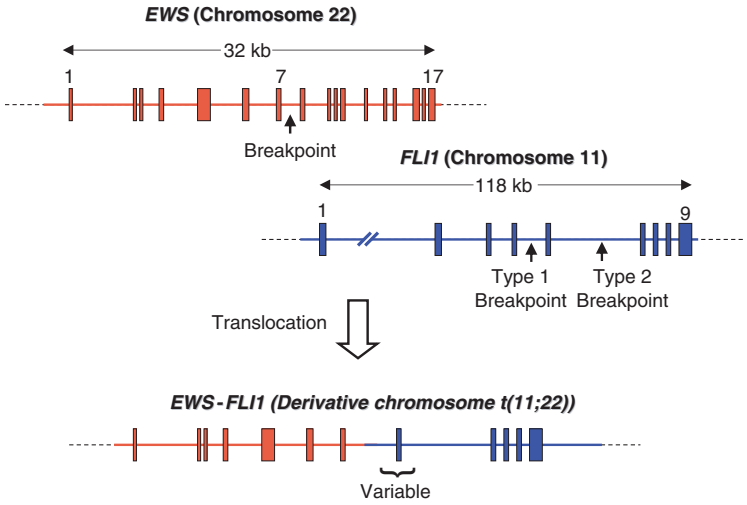

Fig. 2.10 The creation of *EWS-FLI1* by translocation. The *EWS* locus on chromosome 22 spans roughly 32 kb and is composed of 17 exons. In patients with Ewing's sarcoma, a recurring break point is found the seventh intron. The *FLI1* locus on chromosome 11 spans 118 kb and has 9 exons. Within this gene are two recurrent disease-associated break points. In the t(9;22) derivative, the *EWS* and *FLI1* genes are fused, and contain a single open reading frame. The Type 1 and Type 2 fusions result in two distinct *EWS-FLI1* genes that differ in the inclusion of one *FLI1*-derived exon

In all t(11;22) break points, the RNA-binding domain encoded by *EWS* is replaced with the DNA-binding domain encoded by *FLI1*. The *EWS-FLI1*-encoded fusion protein is thus a chimera. Though the target sequences recognized by the DNA-binding domain of the *EWS/FLI1* gene product are indistinguishable from those recognized by native *FLT1*, the chimeric protein is more active and is found to transactivate 5–10 times more transcription than native *FLI1*. Of direct clinical relevance are functional differences between the alternative forms of *EWS-FLI1*. The protein product of the Type 1 fusion was found to be a less effective transcriptional transactivator than the other fusion gene products. This difference in activity correlates closely with the more benign clinical course associated with this alteration.

The *EWS-FLI1* fusion is the most common gene product of chromosomal translocation in Ewing's sarcoma, occurring in about 95% of cases. These alterations are also found in rare tumors that are similar to Ewing's sarcoma. The *EWS* gene has also been found to be fused with several other members of the ETS family of transcription factors in both Ewing's sarcoma and in related disorders.

The discovery of these molecular similarities has led to the reclassification of a group of molecularly and clinically related diseases, which is now referred to as the Ewing's sarcoma-related family of tumors. Cumulatively, these molecular data

suggest that the dysregulation of ETS-mediated transcription by *EWS* fusion is a critical step in the clonal evolution of the Ewing's sarcoma family from their stem cell progenitors.

Oncogene Discovery in the Genomic Era: Mutations in *PIK3CA*

The identification of the majority of known oncogenes predated the sequencing of the human genome. The prototypical oncogenes described in previous sections were isolated on the basis of their homology to genes carried by oncogenic retroviruses or on their ability to induce colony formation in an in vitro transformation assay. These early oncogenes were not discovered because they were necessarily involved in large numbers of cancers. Rather, they emerged as a consequence of idiosyncratic properties that facilitated their discovery by the tools available at the time. While these discoveries were indeed groundbreaking in that they provided a paradigm for understanding how genes cause cancer, the actual genes that emerged were not necessarily contributory to a significant number of cancers. For example, studies of *SRC* genes provided the first critical link between tumorigenic retroviruses and the activation of host cell genes. Yet, mutational activation of *C-SRC* does not appear to contribute to a large proportion of any type of human cancer.

The release of the first draft of the human genome sequence in 2000 has provided a new and powerful means of interrogating the genome of cancer cells. The location, structure and DNA sequence of every human gene is now readily accessible. This information, combined with incremental improvements in DNA sequencing technology, has facilitated the direct analysis of the genes that are mutated in cancer cells. Oncogene discovery is now a systematic process that relies heavily on informatics, the study and processing of large volumes of complex information. In the genomic era, new oncogenes are discovered not on the basis of an idiosyncrasy or serendipity, but on the basis of their frequency of mutation in cancers.

An example of an oncogene identified by cancer genomics is *PIK3CA*. *PIK3CA* is a member of a family of genes that encode lipid kinases known as *phosphatidylinositol 3'-kinases* (PI3Ks). The PI3K enzymes first became a focus of interest to cancer researchers in the 1980s, when it was found that PI3K activity was linked to the protein products of viral oncogenes, such as *C-SRC*. PI3K enzymes function in the signaling pathways involved in tissue homeostasis, including cell proliferation, cell death, and cell motility. The organization of these signaling pathways and the role of PI3Ks in cancer phenotypes will be described in detail in Chapter 5.

The known roles of the lipid kinases in cancer-associated cellular processes and their association with known viral oncogenes formed the rationale for the large-scale analysis of all genes in this family. As part of an attempt to scour the genome for cancer genes, a group at Johns Hopkins University led by Victor Velculescu used informatics to identify eight members of the PI3K family, all related by similarities in their coding sequences. Each of the PI3K genes identified contained a putative kinase domain at its C-terminus. The Johns Hopkins group proceeded to sequence

Table 2.3 Activating mutations in *PIK3CA*

Cancer type	Mutation frequency (%)
Breast	40
Endometrial	36
Colorectal	32
Gastric	25
Ovarian	2–7*
Brain	3–27*
Lung	4

*Varies depending on cell type of origin.

the 117 exons that, in total, encoded the kinase domains of each of the PI3K-family members in a panel of colorectal tumors. Recurrent mutations were found in a single family member, *PIK3CA*. Expanding their analysis to include all *PIK3CA* coding exons in nearly 200 tumor samples, the Johns Hopkins group established that *PIK3CA* is mutated in 32% of colorectal cancers.

The majority of mutations that occur in *PIK3CA* during colorectal tumorigenesis are single nucleotide substitutions that result in missense mutations. These mutations do not occur at random points along the *PIK3CA* open reading frame, but rather occur in clusters known as *hot spots*. Most frequently mutated was a helical domain that largely defines the three dimensional structure of the encoded protein. Also frequently mutated was the C-terminus portion of the lipid kinase domain. The amino acid residues that are affected by hot spot mutations are highly conserved among evolutionarily-related proteins. Functional studies of *PIK3CA* mutants have shown that hot spot mutations cause an increase in the enzymatic activity of the encoded protein.

These initial sequencing efforts also revealed mutations of *PIK3CA* in brain tumors, and breast, lung and gastric cancers (see Table 2.3). Subsequent analysis of additional cancer types has shown that *PIK3CA* is mutated in a large proportion of endometrial cancers, which arise in the epithelial lining of the uterus, and ovarian cancers. Overall, the mutated alleles of *PIK3CA* are among the most prevalent of all cancer genes.

Selection of Tumor-Associated Mutations

Mutations identified via high throughput approaches are not identified on the basis of their function, but on the basis of their sequence. Because of a high degree of sensitivity and specificity, genomic DNA sequencing can reveal all base changes, passenger mutations and driver mutations alike (see Chapter 1). From a practical standpoint how can a geneticist discriminate between a passenger mutation that occurred at random and a driver mutation that actually contributed to tumorigenesis? The evaluation of the *PIK3CA* mutations found in cancers provides a worked example of how careful analysis can discriminate passengers and drivers.

What is the evidence that *PIK3CA* is a cancer gene and not simply a target of passenger mutations? The first and strongest piece of evidence is the large number and frequency of *PIK3CA* mutations that are found in many different tumor samples. Passenger mutations are clonally expanded by chance and are thus predicted to be rare. The observed mutations in *PIK3CA* hot spots were found to occur at a rate that was over 100-fold above the background rate of nonfunctional alterations that had previously been observed in colorectal cancer cells. In contrast, high throughput sequencing of other genes, such as the other members of the PI3K family, has revealed low levels of base changes that are consistent with passenger mutations.

A second piece of evidence is the proportion of silent mutations to missense mutations observed. Silent mutations, which in this context are referred to as *synonymous mutations*, should confer no selective advantage because by definition such mutations do not result in changes to the encoded protein. Missense mutations, or *nonsynoymous mutations*, potentially confer a selectable advantage. Among mutations that are propagated by chance alone, nonsynonymous mutations would be expected to occur at a rate that is about twofold the rate of synonymous mutations. This is simply a function of the numbers of potential bases changes that can occur at random within an open reading frame. The nonsynonymous mutations found in the *PIK3CA* gene occur at a frequency 30 times higher than synonymous mutations in the same gene. This overrepresentation of nonsynonymous mutations indicates a high probability that they conferred a selective advantage, and thus contributed to tumorigenesis.

The clustering of mutations in evolutionarily conserved hot spots of *PIK3CA* is also significant. As described in Chapter 1, evolutionarily conserved protein elements tend to be fundamental to protein function. Therefore, the frequency of mutation at these key codons provides another convincing piece of evidence that the mutations observed in colorectal tumors are highly likely to confer functional phenotypic changes that, in turn, promote cancer cell growth.

Multiple Modes of Proto-oncogene Activation

There are several ways in which changes to the genome can result in the activation of proto-oncogenes. Whether a mutation results from a small sequence alteration, gene amplification, a chromosomal translocation or another more complex gross rearrangement, the contribution of an oncogene to tumorigenesis is qualitatively the same. Somatic mutations that activate proto-oncogenes increase the activity of the encoded protein.

Increased protein activity can result from increased levels of gene expression, as we have seen in the examples of the commonly amplified *MYC* and *ERBB2* oncogenes and in the cases in which *C-MYC* is relocated to a position upstream of a highly active promoter. Alternatively, somatic mutations can result in the expression of a mutant protein. In the case of the *RAS* gene family, activating point mutations result in a loss of regulation and constitutive enzymatic activity. In the case of the

more complex *BCR-ABL* and *EWS-FLI1* oncogenes, the fusion of unrelated genes results in both a change in transcriptional activation and a dramatic change in protein structure. Both of these factors can contribute to increased activity of oncogenic proteins.

Another general theme that emerges from a survey of commonly activated oncogenes is that the same oncogene can be activated by different kinds of mutations in different cancers. As we have seen, *C-MYC* is activated by amplification in a significant proportion of breast and ovarian cancers, but activated by rearrangements in Burkitt lymphoma and in B-cell and T-cell leukemias. The mechanism of activation in a single cancer type is not always exclusive. While *ERBB2* is most frequently activated by amplification in breast cancers (20%), nonsynonymous single nucleotide substitutions are found in lower levels in breast (4%) and also in ovarian (10%), gastric (5%) and colorectal (3%) cancers. Similarly, *PIK3CA* is activated by single nucleotide substitutions in a wide range of carcinomas. While single nucleotide substitutions within *PIK3CA* mutations are found in a small proportion of high-grade ovarian carcinomas, about 15% of such tumors harbor amplifications of this locus.

In a few cases, the causal relationship between a cancer type and a specific mechanism of proto-oncogene activation is fairly obvious. In the cellular precursors of many leukemias and lymphomas, for example, specific immune response genes are transcriptionally much more active than in any other cell type. It is easy to imagine that any chromsomsomal event that results in the juxtaposition of a growth promoting gene such as *C-MYC* with a highly active gene would result in a strong selectable advantage, and outgrowth of that clone.

In most types of cancers, the reason for an apparent bias towards the activation of a proto-oncogene by one mechanism versus another is unclear. One important factor, to be discussed in more detail in Chapter 4, is that different types of cancer cells are inherently prone to different kinds of genomic alterations. Some cancers are characterized by gross numerical and/or structural chromosomal abnormalities, while others exhibit a preponderance of changes that occur at the nucleotide level. The acquisition of different forms of genetic instability during tumorigenesis is an important factor in determining the spectrum of somatic mutations present in an advanced cancer.

Oncogenes are Dominant Cancer Genes

As we have seen throughout this chapter, a single mutation is sufficient to activate a proto-oncogene and convert it to an oncogene. The activating mutation results in a growth advantage, in spite of the continued presence of a normal, unmutated allele in every cell. Because the phenotype conferred by an oncogenic mutation is not masked by the presence of the remaining wild type allele, oncogenes are, by definition, dominant alleles.

The oncogenic mutations found in a tumor sample are almost never found in the normal cells of that same individual. The few known exceptions to this pattern are

described in the following section. Generally, an activated oncogene is not found in the germline of a cancer-prone family. Extensive examination of proto-oncogenes and oncogenes in normal tissues and in cancers has revealed that the nearly all of the mutations that convert proto-oncogenes to oncogenes are acquired by somatic mutation.

Cancer genes can be acquired by somatic mutation or by inheritance. Cancer predisposition is an inherited trait, and therefore the genes that confer this trait must be present in the germline. Oncogenes are not commonly found in the germline and therefore probably are not a major factor in cancer predisposition. Clearly, this is true for oncogenes that are highly *penetrant*, that is, those that exert strong phenotypic effects regardless of environment or genetic background. It remains to be revealed whether less penetrant oncogenes, with relatively subtle phenotypes, will play a significant role in the inheritance of cancer. Types of cancer that are clearly heritable are attributable to another type of cancer gene entirely: the tumor suppressor gene. The nature of these important cancer genes will be described in chapter 3.

Germline Mutations in *RET* and *MET* Confer Cancer Predisposition

All of the oncogenes described thus far are activated by somatic mutations that occur during tumorigenesis. An interesting exception to this general pattern is provided by the *RET* oncogene, which is somatically mutated in cancers, but is also found in the germline of individuals that are predisposed to inherited cancers of the endocrine system.

Multiple endocrine neoplasia type 2 (MEN2) is a rare, autosomal dominant cancer syndrome. There are several clinically distinct subtypes of this inherited disorder, designated MEN2A, MEN2B and familial medullary thyroid carcinoma (FMTC). Affected individuals most commonly develop an atypical form of thyroid carcinoma which is derived from a population of cells that have an origin in the neural crest. Other endocrine cancers, benign lesions and developmental abnormalities are variably seen in the different MEN2 subtypes.

MEN2-related cancers are caused by germline mutations in the *RET* proto-oncogene. The *RET* proto-oncogene is located on chromosome 10 and contains 21 exons that encode a membrane-bound tyrosine kinase. Like many other oncogenes, *RET* was first discovered during *in vitro* transformation assays using genomic DNA from lymphomas and gastric tumors. It was found that the first isolates of this gene were actually chimeras that had formed during the transfection process. The gene was accordingly designated by the acronym for '*re*arranged during *t*ransfection'. The oncogenic forms of *RET* that have been found in sporadic cancers are similarly rearrangements. These somatic rearrangements vary in different cancers, but commonly put the tyrosine kinase domain in frame with highly expressed genes, thereby resulting in its constitutive activation. These types of mutations are different from those that cause MEN2.

In contrast to the *RET* mutations found in sporadic cancers, the mutations found in individuals affected with MEN2 are usually single nucleotide substitutions. Activating point mutations that convert *RET* into an oncogene typically affect the extracellular domain of the *RET*-encoded protein and lead to ligand-independent activation of the kinase and constitutive activation of downstream mitogenic pathways. (These pathways and the manner in which they related to cancer cell phenotypes will be described in Chapter 5.) Most commonly, mutations in *RET* affect exons 8 and exons 10–16. The precise location of the mutations is associated with distinct disease phenotypes.

RET is one of a small number of oncogenes that causes an inherited predisposition to cancer. Another oncogene known as *MET* is carried in families affected by hereditary renal cell carcinoma. Like the MEN2 syndomes, hereditary renal cell carcinoma is rare, but highly illustrative of the role that oncogenes can play in some inherited forms of cancer.

The role of the oncogenic forms of *RET* and *MET* in heritable cancers is highly unusual. Activated oncogenes are dominant alleles. As will be extensively discussed in Chapter 3, the cancer genes that contribute to hereditary forms of cancer are almost always recessive alleles that are unmasked during the process of tumorigenesis. While it is entirely possible that genomic analysis of cancer-prone families will uncover additional germline oncogenes, it is clear that penetrant, dominant cancer genes would strongly disfavor the viability of carriers. The role of oncogenes in inherited cancer predisposition appears at this point to be relatively small.

Proto-oncogene Activation and Tumorigenesis

How do oncogenes fit into the sequence of genetic alterations that underlie tumorigenesis? The oncogenes that contribute to colorectal tumorigenesis are highly informative (see Fig. 2.11). The genetic changes that most frequently occur in these tumors can be directly associated with discrete clinico-pathological stages of the disease. The activation of oncogenes is seen in most, if not all, colorectal cancers. There are several oncogenes that are frequently found, and these have been shown to be stage-specific.

The first cancer genes firmly associated with colorectal cancers were activated members of the RAS family. Single nucleotide substitutions within *K-RAS* and *N-RAS* are found in approximately 50 percent of all colorectal cancers. Among the precancerous lesions, adenomas greater than 1 cm in size exhibit a frequency of *RAS* mutations that is similar to that seen in invasive cancers. In contrast, smaller adenomas (<1 cm) rarely exhibit *RAS* mutations. This finding suggests that *RAS* mutations are acquired during adenoma progression. In support of this hypothesis, dissection of adenomas has revealed small subpopulations in which *RAS* mutations have occurred. Presumably, these subpopulations represent clones that are beginning to progress and that have the potential to give rise to more invasive clones.

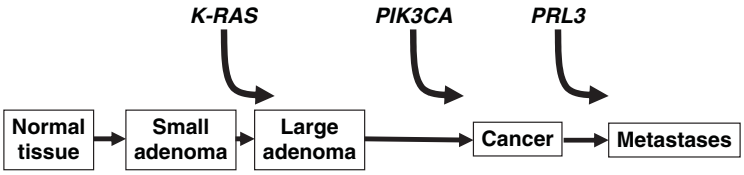

Fig. 2.11 Oncogenes and colorectal cancer progression. Oncogenic mutations in *K-RAS, PIK3CA* and *PRL3* contribute most clearly to later stages of cancer progression. While *K-RAS* mutations are occasionally found in very small aberrant crypt foci (ACF), ACF that harbor *K-RAS* mutations do not appear to progress. In contrast, *K-RAS* activation plays an important role in the transition from small to large adenomas, which have significant potential to become malignant. *PIK3CA* mutations are largely restricted to invasive cancers, while *PRL3* is amplified in a significant proportion of metastatic lesions

Interestingly, *RAS* mutations can be found in some very early lesions arising in the colorectal mucosae. In a distinct histological subset of the earliest lesions, the aberrant crypt foci, *RAS* mutations are found at a high rate. However, such lesions are self-limited and appear to have little, if any, potential for progression. This is a very illuminating finding that underscores a basic principle of colorectal tumorigenesis. Clearly, *RAS* mutations can occur in any cell population, but they alone are not sufficient to promote the continued growth of a neoplasm. Rather, the stepwise expansion of tumor cell clones requires a defined sequence of events. While *RAS* mutations appear to be of primary importance in the progression of adenomas to more advanced tumors, this effect is stage-specific and requires prior genetic alterations.

Mutational activation of *PIK3CA* also occurs frequently in colorectal cancers. As with *RAS* mutations, these mutations are not found in early-stage tumors. Rather, *PIK3CA* mutations usually arise late in the process of tumorigenesis, close to the point that a tumor begins to invade surrounding normal tissues. Based on this finding, it appears likely that increased activity of the *PIK3CA*-encoded protein provides a survival advantage to cancer cells as they penetrate the barriers that physically separate tissue compartments.

The most deadly phase of tumorigenesis begins when cancer cells metastasize and colonize distant tissues. This is the most clinically intractable phase of the disease but, unfortunately, the phase that is least understood at the genetic level. In the case of advanced colorectal cancers, metastatic tumors most frequently arise in the liver. Studies of global gene expression have revealed a gene, *PRL-3* that is activated nearly all liver metastases but rarely in primary colorectal tumors. *PRL-3* encodes a tyrosine phosphatase, an enzyme that removes phosphate moieties from protein tyrosine residues. Functional studies of this enzyme have shown that it can promote cell migration, invasion and metastasis in experimental systems. Genetic analysis of the *PRL-3* locus has revealed increased copy number and gene amplifications in 45% of metastatic lesions. Thus, in the majority of cases, the mechanism of *PRL-3* activation is unknown.

Double minute chromosomes have been observed in a significant fraction of colorectal tumors, suggesting that gene amplification is a frequent occurrence. However, the proto-oncogenes that may be activated within these amplicons, and their role in tumorigenesis remains poorly defined. Several genes, including *C-MYC* and *ERBB2* have been found to be amplified in small numbers of colorectal tumors. Unlike *RAS, PIK3CA* and *PRL-3*, these oncogenes have not been found to predominate in any specific stage of cancer and their potential roles in the process of tumorigenesis remain uncertain.

Further Reading

Bishop, J. M. Enemies within: The genesis of retrovirus oncogenes. *Cell* **23**, 5–6 (1981).

Epstein, M. A. Historical background. *Philos. Trans. R. Soc. Lond. B. Biol. Sci*. **356**, 413–420 (2001).

Garraway, L. A. & Sellers, W. R. Lineage dependency and lineage-survival oncogenes in human cancer. *Nat. Rev. Cancer* **6**, 593–602 (2006).

Martin, G. S. The road to Src. *Oncogene* **23**, 7910–7917 (2004).

Mitelman, F., Johansson, B. & Mertens, F. The impact of translocations and gene fusions on cancer causation. *Nat. Rev. Cancer* **7**, 233–245 (2007).

Rowley, J. D. Chromosome translocations: Dangerous liaisons revisited. *Nat. Rev. Cancer* **1**, 245–250 (2001).

Sawyers, C. L. Chronic myeloid leukemia. *N. Engl. J. Med*. **340**, 1330–1340 (1999).

Schwab, M. Oncogene amplification in solid tumors. *Semin. Cancer Biol*. **9**, 319–325 (1999).

Thomas, R. K. et al. High-throughput oncogene mutation profiling in human cancer. *Nat. Genet*. **39**, 347–351 (2007).

Weinberg, R. A. The cat and mouse games that genes, viruses, and cells play. *Cell* **88**, 573–575 (1997).

Chapter 3
Tumor Suppressor Genes

What is a Tumor Suppressor Gene?

A *tumor suppressor gene* is a type of cancer gene that is created by loss-of-function mutations. In contrast to the activating mutations that generate oncogenic alleles from proto-oncogene precursors, tumor suppressor genes, and the proteins they encode, are functionally *inactivated* by mutations.

Tumor suppressor genes typically control processes fundamental to the maintenance of stable tissue compartments. These processes include the maintenance of genetic integrity, the progression of the cell cycle, differentiation, cell–cell interactions, and apoptosis. Mutational inactivation of tumor suppressor genes contributes to the loss of tissue homeostasis – the hallmark of a developing neoplasm.

As described in Chapter 2, the mutations that convert proto-oncogenes to oncogenes are single nucleotide substitutions, amplifications, gene fusions and other chromosomal rearrangements that increase the activity of the proto-oncogene-encoded proteins. Nonsense, frameshift and splice site mutations generally do not lead to proto-oncogene activation. In contrast, tumor suppressor genes are inactivated by mutations; a distinct spectrum of mutations creates this opposite effect. The open reading frames of tumor suppressor genes are commonly truncated by nonsense mutations, small insertions or deletions, or splice site mutations. Larger deletions can eliminate exons or even result in the loss an entire gene. Tumor suppressor genes can also be inactivated by single nucleotide substitutions that alter key residues in the encoded tumor suppressor protein, thereby causing a loss of function in the encoded protein. Gene amplification and chromosomal translocation are generally not mechanisms of gene inactivation.

The Discovery of Recessive Cancer Phenotypes

The first cancer genes to be discovered were oncogenes. For a time it was widely believed that the cancer phenotype resulted primarily from activating mutations that led to gains of function. An early piece of evidence that other types of genetic alteration might also be important in cancer was provided by Henry Harris and his colleagues.

F. Bunz, *Principles of Cancer Genetics*.

© Springer 2008

In a 1969 study, Harris adopted a novel approach to study the genetic factors that were involved in cancer cell phenotypes (see Fig. 3.1). Previously, it had been established that cells of different types could be fused, and thereby made to share their genes. Among the phenotypes that many cultured cancer cells share is an ability to grow into tumors when implanted into mice, a property known as *tumorigenicity*. Tumorigenicity is an experimental trait that is believed to be reflective of the malignant nature of the tumor from which the cultured cancer cell was derived. Harris found that when tumorigenic cells derived from a murine tumor were fused with non-tumorigeneic cells, the hybrids were non-tumorigeneic. The genome of the non-tumorigeneic cells therefore suppressed tumorigenicity of the cancer cells in a dominant manner.

Continued observation of the fused hybrids revealed the basis for this effect. When the hybrid cells were cultured for extensive periods, tumorigeneic subclones within the population began to appear. Cytogenetic analysis revealed chromosomal losses in these revertants. Subsequent studies with human cells demonstrated that the suppression of tumorigenicity was sustained as long as both sets of parental chromosomes were retained. The transfer of individual chromosomes was found to similarly suppress the tumorigenicity of human cancer cells, even if an activated oncogene such as a mutant RAS gene was expressed in the hybrids. Thus, it was apparent that underlying cause of reversion was not chromosomal losses in general,

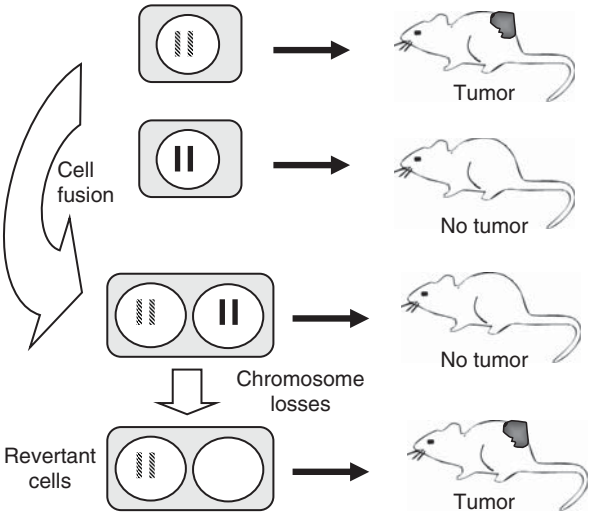

Fig. 3.1 Tumor suppression is a dominant phenotype. Two cell types are isolated from tumors: tumorigenic cells (defined as those that form tumors when introduced into the skin of mice) and non-tumorigenic cells. In this simplified illustration, only the relevant pair of homologous chromosomes is shown in each cell (solid and hatched, respectively). Fusion of these two cells allows them to share their genetic material. Cells containing both sets of chromosomes are not tumorigenic, demonstrating that the alleles that cause tumor formation (carried on the hatched chromosomes) are recessive. Because the chromosome complement of the fused cells are unstable, with time cells appear that have lost wild type alleles (carried on the solid chromosomes) contributed by the non-tumorigenic cells. These rare cells revert to a tumorigenic phenotype

but rather the loss of specific chromosomes. It was concluded that tumorigenicity was a recessive trait that could be suppressed by the transfer of a specific chromosome and perhaps even by the transfer of a single dominant gene.

Somatic cell genetic studies such as these supported the idea that at least some aspects of the malignant cancer cell are recessive traits that arise through genetic losses. It is important to understand that the primary assay used in these studies, the generation of tumors upon introduction of cells into a mouse, does not recapitulate the many selective pressures faced during the evolution of a naturally occurring tumor. Thus, the chromosome additions and subsequent losses by the hybrids presumably affected a subset of cancer-associated phenotypes. Indeed, other cancer cell phenotypes such as immortality and anchorage independent growth were found to be retained in the somatic cell hybrids. These results are consistent with the idea that cancer cells arise by the accumulation of multiple genetic alterations, causing the sequential acquisition of stage-specific phenotypic traits. The genetic alterations that promote tumorigenesis cause both the gain of oncogene function as well as the loss of tumor suppressor genes.

Retinoblastoma and Knudson's Two-Hit Hypothesis

Retinoblastoma is a tumor of the eye that arises from immature cells within the retina (see Fig. 3.2). Though retinoblastoma is a relatively rare tumor, occurring with an incidence of roughly 1 in 25,000, its clinical features and the genetic

Fig. 3.2 Retinoblastoma. A pediatric malignancy of the retina, retinoblastoma occurs in hereditary and sporadic forms. Shown is a patient with a unilateral tumor (in the eye on the left). (Courtesy of the National Cancer Institute.)

factors that influence its incidence have provided a model for understanding how tumor suppressor genes work.

Retinoblastoma occurs during infancy and early childhood. The cells of the retina are nearly fully differentiated at birth and have a limited capacity for further proliferation. This limitation provides, in effect, a window of susceptibility. Somatic mutations that occur within this window can affect a proliferating cell population from which cancers can arise.

Retinoblastoma has distinctive clinical features that allow it to be diagnosed with a high degree of accuracy. These cancers occur in both sporadic and inherited forms. Examination of epidemiological data by Alfred Knudson in the early 1970s revealed that inherited retinoblastomas that occur in individuals with a family history of the disease frequently affect both eyes and exhibit multifocal tumors in a single eye. In contrast, the sporadic form of retinoblastoma, which occurs in the absence of a family history, is always unifocal and restricted to one eye. In addition, Knudson noted that the bilateral and multifocal cases have an earlier age of onset.

Building upon these data, Knudson formulated what would come to be known as the two-hit hypothesis (see Fig. 3.3). Knudson deduced that two genetic alterations, or 'hits', are necessary for retinoblastoma development. In individuals with the inherited form of the disease, the first hit is acquired via the germline and thus present in every cell. The second hit required for the disease in these predisposed individuals is a somatic mutation in a single cell that would then expand into a tumor. In sporadic cases of retinoblastoma, both of the required hits were acquired via somatic mutations by a precursor cell.

The Knudson model explains the earlier age of onset and the prevalence of bilateral and multifocal tumors that occur in the inherited form of the disease. The rate limiting event is the somatic mutation, which occurs at random and at a very low frequency during the limited window of susceptibility. Individuals predisposed to retinoblastoma already carry the first hit in every cell of their body. The onset of disease thus requires only one rate-limiting step. In sporadic cases, two rare events occur within a single cell within a brief window of time. Sporadic retinoblastoma is accordingly rare, has a later age of onset and presents as a single lesion in one eye.

The formulation of the two-hit hypothesis by Knudson was contemporary with the somatic cell genetic studies undertaken by Harris. These two very disparate lines of inquiry converged on a common principle: recessive genetic determinants play a significant role in cancer. While these studies laid the framework for understanding the roles of tumor suppressor genes, the actual loci that possessed these properties remained to be identified. Only upon the actual cloning of tumor suppressor genes did the extent of their involvement in cancer become apparent.

Chromosomal Localization of the Retinoblastoma Gene

Based on epidemiology alone, the two hits required for retinoblastoma could have been accounted for by mutations in two separate genes, perhaps two oncogenes that were sequentially activated. However, the experiments by Harris had shed light on

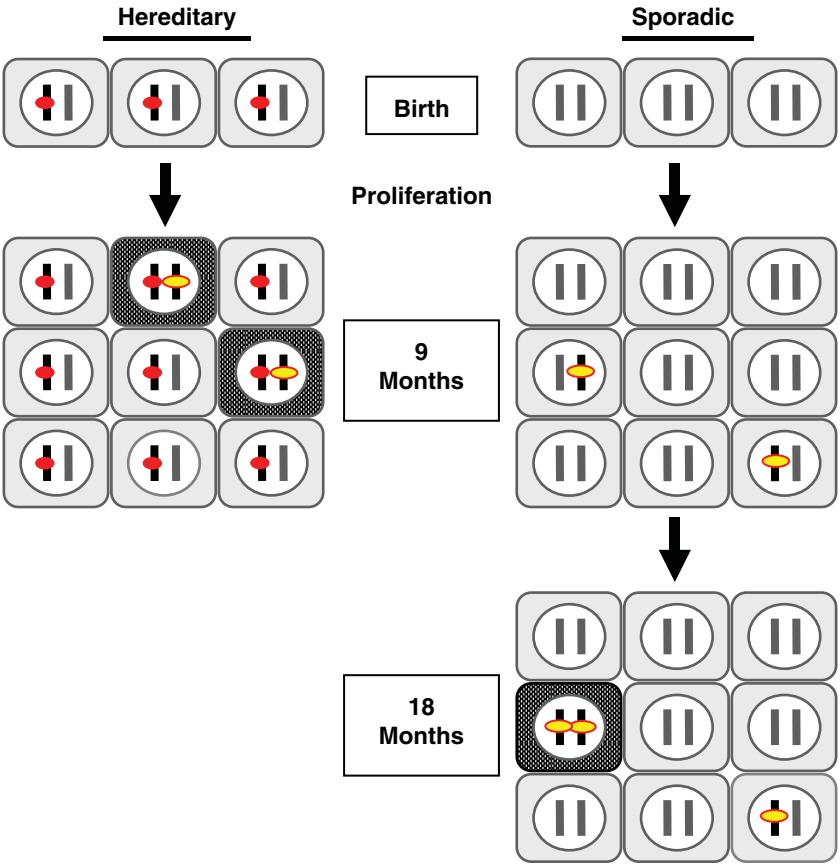

Fig. 3.3 The 'two-hit' hypothesis. At birth, individuals predisposed to the hereditary form of retinoblastoma harbor a mutant allele (shown in red) in every cell, including the blast cells of the retina. These cells proliferate during the first 9 months of life. During this time, somatic mutations at the retinoblastoma locus (shown in yellow) occur at a low frequency. In individuals predisposed to retinoblastoma, the somatic inactivation of a single allele is sufficient to provide the two hits required for tumor formation. Multiple tumor precursor cells (shown as dark cells) are thus generated, leading to bilateral tumors that are often multifocal. In contrast, normal cells require two somatic hits for tumor development – a low probability event. Because of the requirement for two somatic hits, sporadic retinoblastomas are rare and tend to occur at an older age than inherited retinoblastomas

a completely novel and uncharacterized type of cancer gene that was recessive in nature. Knudson suggested that both hits could occur in the two alleles of a single locus, and thereby cause the total inactivation of a recessive cancer gene. Knudson's hypothesis was supported and extended with the cloning and identification of the retinoblastoma susceptibility gene, *RB*.

The genetic basis of retinoblastoma was inferred from its distribution among sporadic and familial cases. However, the identity of the actual gene or genes

that might be the target of Knudson's hits was completely unknown. The progressive mapping and ultimately the identification of the retinoblastoma gene was a landmark effort that not only determined the molecular basis of retinoblastoma susceptibility, but also revealed a general and central principle of cancer genetics.

Retinoblastoma tumor cells had no known biochemical or signaling defect. There was therefore no basis to suspect that the causative gene might encode any particular enzyme or regulatory protein. Ultimately, the retinoblastoma gene was not identified by virtue of its function, but rather on the basis of its location.

The first clue as to the location of the retinoblastoma locus arose from cytogenetic studies in the late 1970s. Improvements in chromosome banding techniques facilitated the detailed analysis of karyotypes from normal and tumor-derived cells. Microscopically visible deletions within one copy of chromosome 13 were observed in the normal blood cells of a small proportion of individuals with the inherited form of the disease. A cytogenetic aberration that is found in the normal cells of an individual is known as a *constitutional* alteration.

While the extent of the constitutional deletion varied among these individuals, one particular chromosomal band on the short arm of chromosome 13, designated 13q14, was consistently missing. Somatic deletions involving band 13q14 were also found in about 25% of sporadic retinoblastoma tumor samples. In these cases, examination of blood cells from the same patient revealed two normal homologs of chromosome 13, indicating that the 13q14 deletions found in the tumors were somatic mutations that occurred during tumor development. Thus, a defined chromosomal deletion within the body of a chromosome, known as an *interstitial* deletion, appeared to be associated with at least some cases of both inherited and sporadic retinoblastoma.

The pattern of these interstitial deletions in normal and cancer cells was consistent with the two-hit model proposed by Knudson. The constitutional deletion found in patients with inherited retinoblastoma was present in all cells and consistent with the first hit. The deletions present in the cancer cells obtained from the sporadic cases were clearly somatic hits that had occurred during tumorigenesis, though it was not yet clear that they had initiated the process.

The recurrent nature of the 13q14 interstitial deletion suggested that a gene within the 13q14 band might be the target of the two hits required for retinoblastoma development. However, in the majority of retinoblastoma cases, no obvious karyotypic abnormalities were seen. It remained a possibility that a smaller deletion, undetectable by cytogenetic approaches, might constitute a genetically equivalent hit in these cases. Deletions that can be recognized by the observation of whole chromosomes are large, typically extending over regions that can span hundreds of kilobases. Additional evidence was needed in order to establish whether the retinoblastoma susceptibility locus was located within the cytogenetically defined 13q14 region.

The mapping of the retinoblastoma susceptibility gene to the13q14 region was confirmed by linkage analysis, using a nearby gene as a point of reference. The *ESD* gene, which encodes an enzyme known as esterase D, had recently

been localized to the 13q14 band by William Benedict and his coworkers. While the biochemical function of esterase D was irrelevant to the pathogenesis of retinoblastoma, *ESD* could be used as a genetic marker to track 13q14 deletions in individuals that did not exhibit gross karyotypic abnormalities.

Esterase D exists in two distinct forms that are encoded by two common *ESD* alleles. The two forms of esterase D protein could be resolved by protein electophoresis, and thus heterozygous individuals could be identified. Heterozygosity at the *ESD* locus allowed the tracking of each allele through the pedigrees of families with inherited retinoblastoma. William Benedict and his coworkers determined that, within a disease-prone family, children that inherited one allele of *ESD* invariably developed retinoblastoma while children that inherited the other *ESD* allele remained disease-free (see Fig. 3.4). Put in genetic terms, the *ESD* alleles cosegregated with the retinoblastoma susceptibility trait. These studies revealed that *ESD* and the as-yet undiscovered retinoblastoma gene were tightly linked and therefore in close physical proximity to one another.

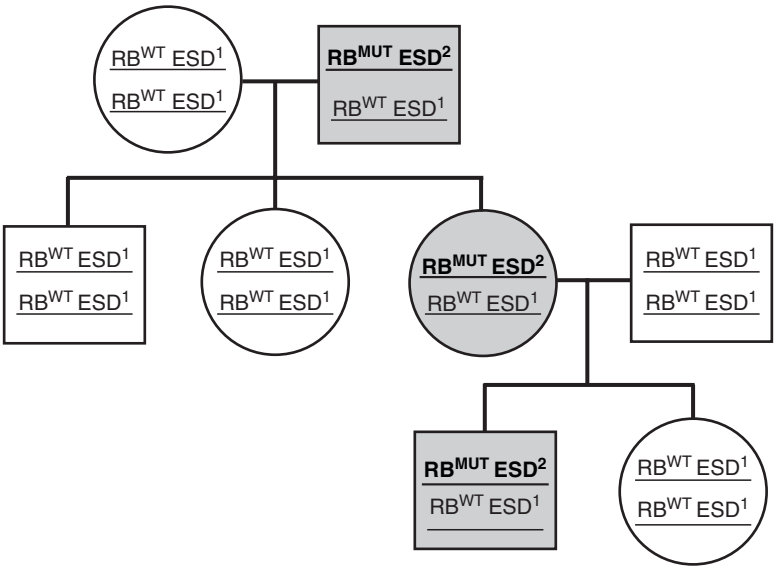

Fig. 3.4 Linkage between the putative *RB* locus and *ESD*. Distinct alleles of the *ESD* gene (denoted ESD^1 and ESD^2), encode proteins that can be resolved by protein electrophoresis. Evaluation of *ESD*-encoded proteins thus provides an assessment of *ESD* allelotype. In retinoblastoma kindreds, allelic variants of *ESD*, when present, invariably cosegregate with disease. In the pedigree shown, circles represent females, squares represent males. Shaded circles and squares represent individuals affected with retinoblastoma

The Mapping and Cloning of the Retinoblastoma Gene

The linkage between the retinoblastoma susceptibility locus and *ESD* provided a means of exploring the nature of the hits that involved a putative retinoblastoma gene in the 13q14 region. Cases were identified in which normal cells of a retinoblastoma patient were heterozygous for *ESD*. Heterozygosity of the nearby *ESD* locus did not affect retinoblastoma predisposition, but rather created a situation in which the two chromosomal loci could be distinguished. Heterozygosity is thus informative; no additional information could be extracted from homozygous cases where the *ESD* alleles were indistinguishable.

In several of these informative cases, both *ESD* alleles were present in normal blood cells, while only one of the two *ESD* genes was present in the cells of the tumor (see Fig. 3.5). The tumor cells thus exhibited a *loss of heterozygosity*, or LOH, in the region adjacent to the susceptibility locus. LOH was also detected in inherited cases of retinoblastoma. In these cases, the allele of *ESD* retained in the tumor cells was invariably the one that was inherited from the affected parent. The reduction of the *ESD* locus to homozygosity was a strong piece of evidence that a somatic mutation resulting in a loss of genetic material had functionally inactivated the single normal copy of the putative retinoblastoma gene in susceptible individuals. Thus, the link between the retinoblastoma locus and *ESD* provided

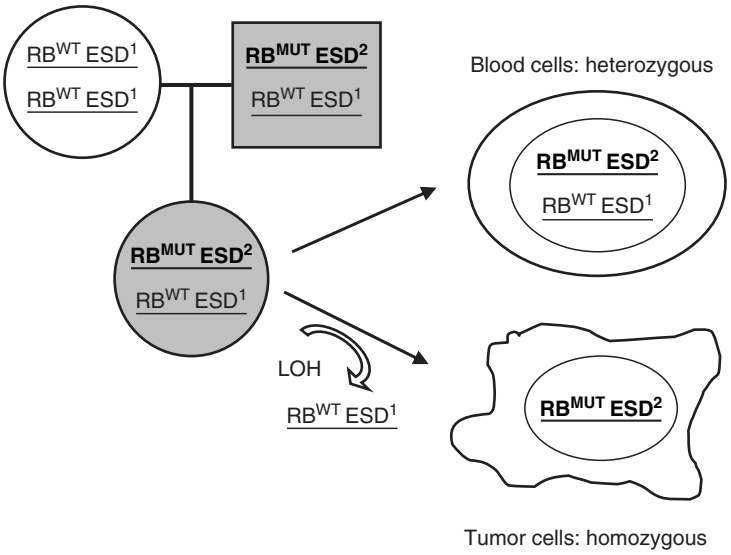

Fig. 3.5 Loss of heterozygosity in the region of the putative *RB* locus. The mutation that causes heritable retinoblastoma (RB^{MUT}) is present in the germline of an affected individual. In this example, RB^{MUT} cosegregates with a distinguishable *ESD* allele, ESD^2. During tumor development, the single normal allele and the ESD^1 allele to which it is linked are invariably lost. Only ESD^2 is detectable in tumor cells. In contrast, both alleles are retained in normal blood cells

an indirect means of assessing genetic losses in that region. While the 13q14 region was lost in only a small fraction of retinoblastoma patients, the LOH at *ESD* suggested that submicroscopic deletions could account for the loss of a neighboring gene in many additional patients.

The precise location and the identity of the actual retinoblastoma gene remained to be determined. The fact that a tumor suppressor locus was strongly linked to a polymorphic gene that encodes readily distinguishable proteins was highly fortuitous. In general, tumor suppressor loci cannot be evaluated for losses by examining expressed proteins. A more versatile mapping strategy is to link unknown genes with known DNA polymorphisms using methods of DNA analysis.

One type of polymorphism that could be detected with the information and technology available during the 1980s was the *restriction fragment length polymorphism*, or RFLP. Restriction enzymes are enzymes that cut DNA at sites that have defined and highly specific recognition sequences. Some DNA sequence polymorphisms result in changes in the pattern at which these recognition sequences occur in a chromosomal region. Differences between homologous chromosomal loci could therefore be detected by examining the lengths of the polymorphic restriction fragments produced by the digestion of genomic DNA. Specific fragments could be visualized by the technique of Southern blotting, in which a defined, locus-specific probe is hybridized to a restriction fragments that are fractionated by length.

In 1983, RFLP analysis performed by Webster Cavanee, Ray White and their colleagues confirmed and extended the linkage between *ESD* and the putative retinoblastoma gene. They were able to track specific heterozygous restriction fragments from affected parents to affected children, and were able to detect LOH in tumor cells. With the use of additional probes, Thaddeus Dryja and his coworkers were able to detect relatively small regions within the 13q14 region that were homozygously deleted in retinoblastoma tumor cells. Thus, two significant milestones had been simultaneously attained. First, the direct relationship between genetic loss and retinoblastoma cancer development was firmly established. Second, the location of the putative retinoblastoma locus had been narrowed down to a relatively short region of the chromosome. The technique of using a series of adjacent probes to systematically examine loss patterns along lengthy regions was a powerful technique that became known as ‘walking the chromosome’.

The retinoblastoma gene, *RB*, was identified in late 1986 and early 1987 by three independent groups. In collaboration with Robert Weinberg and his laboratory, the Dryja group cloned the retinoblastoma gene by hybridizing cloned genomic DNAs from within the known region of loss with fractionated RNA transcripts from normal retinal cells and retinoblastoma cells. A 4.7 kb RNA transcript was identified that was present only in normal cells and not in tumor cells. The laboratory of Wen–Hwa Lee and that of Yuen–Kai Fung and William Benedict also cloned the *RB* gene using the same general strategy. The approach of using localized markers to clone genes came to be known as *positional cloning*.

The cloning and characterization of the *RB* gene facilitated detailed mutational analysis in large tumor panels. *RB* consists of 27 exons that span a genomic region that is approximately 178 kb in length. Subsequent mutational studies detected

frequent deletions that eliminate all or part of the coding region. Other types of mutations that result in *RB* inactivation occur at lower frequencies.

Functional analysis of the *RB* gene has revealed a central role for its encoded protein in the regulation of the cell cycle. Homozygous inactivation of the *RB* gene results in a total loss of RB protein function and a corresponding lack of cell cycle regulation. This loss of regulation is apparently sufficient to initiate tumor initiation in the immature cells of the retina. The role of RB in the regulation of the cell cycle will be described in detail in Chapter 5.

Tumor Suppressor Gene Inactivation: The Second 'Hit' and Loss of Heterozygosity

One of the key observations that guided the discovery of the *RB* gene and confirmed its recessive nature was LOH, the reduction to homozygosity of a locus that previously was heterozygous. LOH is the second 'hit' predicted by Knudson, and represents the loss of the remaining wild type allele of a recessive tumor suppressor gene. With current methods, LOH can readily be assessed by the examination of known single nucleotide polymorphisms (SNPs), which provide convenient and easily detectable genetic markers (see Chapter 1).

During the process of tumorigenesis, LOH can occur via a number of mechanisms (see Fig. 3.6):

Loss of a whole chromosome. Chromosome nondysjunction during mitosis can cause an imbalance in chromosomal segregation, resulting in a chromosome loss in one daughter cell. Nondysjunction can sometimes be followed by reduplication of the remaining unpaired chromosome. In these cases, the overall number of alleles is preserved. Following chromosome loss, heterozygosity of all genes and markers on a chromosome is also lost, regardless of whether reduplication takes place.

Mitotic recombination. Pairing of homologous chromosomes during mitosis results in crossing over and physical exchange of genetic material. Recombination of this type occurs most frequently during meiosis, but also occurs at a lower rate during mitosis. The LOH that results is also known as gene conversion. LOH resulting from mitotic recombination is restricted to a portion of a chromosome.

Localized mutations. The remaining wild type allele of a tumor suppressor gene can be lost by a second mutation that does not involve a large chromosomal region, but is rather more local in nature. In retinoblastomas, for example, deletions have frequently been found to inactivate both *RB* alleles. In such cases, the first and second 'hits' occur via the same mechanism but independently of one another, resulting in a homozygous deletion. Note that other types of mutations can also confer the second 'hit' predicted by the Knudson hypothesis, without necessarily causing LOH.

These processes occur at a higher rate in some cancer cells than they do in normal cells. Correspondingly, the rate of tumor suppressor gene loss is frequently higher in cancer cells than in pre-cancerous precursors. The inherent genetic

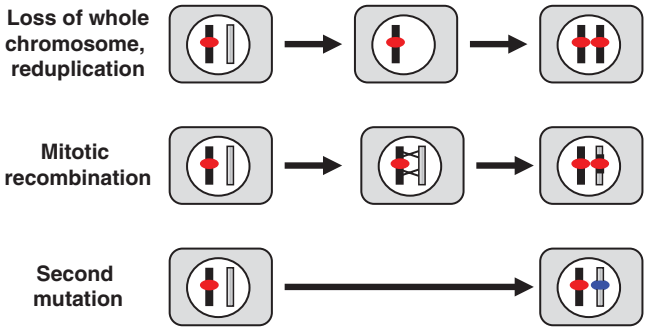

Fig. 3.6 Mechanisms of Loss of Heterozygosity (LOH). A cell contains two homologous chromosomes, one of which contains a genetic alteration (red). LOH can occur by several mechanisms: (1) complete loss of a chromosome, followed by reduplication of the remaining chromosome, will result in LOH at every chromosomal locus (top); (2) recombination during mitosis results in the conversion of a chromosomal region to the homologous region, resulting in a regional LOH (middle); and (3) a second mutation (blue), such as a deletion, can spontaneously arise in the second allele, resulting in a localized region of LOH (bottom)

instability of cancer cells and the contribution of instability to LOH will be described in Chapter 4.

Note that the first two mechanisms of LOH, chromosome loss and mitotitc recombination, involve very large regions of DNA. Therefore, detection of LOH is a very low resolution method of tumor suppressor gene mapping. More local mutations are generally needed to pinpoint a region that contains a putative tumor suppressor locus. In the case of *RB*, it was the discovery of homozygous deletions that provided this crucial information.

In summary, the inactivation of a tumor suppressor gene occurs in two distinct steps that correspond to Knudson's two 'hits'. The first step is the mutational inactivation of one allele. An individual can acquire an inactivated tumor suppressor allele by somatic mutation. Alternatively, an individual can inherit a predisposition to cancer by inheriting a mutated allele via the germline, thereby bypassing a rate-limiting step. This inactivated allele does not, per se, confer any phenotype, but merely causes an increased risk that the gene will subsequently become biallelically inactivated. While the first step varies in inherited and sporadic cancers, the second step to total gene inactivation is always the same: the loss of the remaining normal allele.

Recessive Genes, Dominant Traits

The cloning of the retinoblastoma gene was a landmark in cancer genetics. The incremental localization of *RB* and its linkage to known genetic markers provided a highly successful paradigm for tumor suppressor gene discovery. In recent years,

the sequencing of the human genome and the identification of many defined polymorphisms have greatly streamlined mutational and linkage analyses. As a result, the laborious methods that were used to clone the canonical tumor suppressor genes have become largely obsolete. Nonetheless, the principles uncovered by these early efforts have largely informed our current understanding of how tumor suppressor genes contribute to tumorigenesis and cancer susceptibility.

The identification of *RB* was a considerably more complex and a lengthier process than the isolation of the first oncogenes had been just a decade earlier, despite dramatic improvements in molecular technology. Tumor suppressor genes are inherently more difficult to identify. Tumor suppressor genes are recessive while oncogenes are dominant. It is technically more straightforward to assess the one-step gain of a dominant gene than the two-step loss of a recessive gene. In an experimental setting, a dominant gene such as an oncogene can in many cases recapitulate its cancer phenotype upon introduction into normal cells. In contrast, the cancer-related effects of a mutated tumor suppressor gene are masked by the presence of a normal allele. Recessive phenotypes are expressed only upon the loss of the remaining normal allele.

Individuals are strongly predisposed to retinoblastoma if they inherit a single defective *RB* allele. In general, while tumor suppressor genes are recessive, the inheritance of a mutation in a tumor suppressor gene confers cancer susceptibility, which is a dominant trait. The basis for what, at first glance, seems like a paradox is that recessive cancer genes are unmasked by genetic losses. Such losses are rare. However, there are many cellular targets in which they can occur. In the infant retina, for example, there are more than 10^6 cells, many of which proliferate during the window of susceptibility. Even a rare genetic event is likely to occur in a sufficiently large population of cells. Importantly, tumor suppressor traits such as retinoblastoma are highly penetrant. Thus, even rare events occurring in only a few cells can give rise to multiple tumors.

APC Inactivation in Inherited and Sporadic Colorectal Cancers

Colorectal cancer occurs in both inherited and sporadic forms (see Chapter 1). As in the case of retinoblastoma, the inherited forms of colorectal cancer are caused by germline tumor suppressor gene mutations. Sporadic tumors arise as a result of somatically required mutations in the same gene. In colorectal cancer, a gene that plays a role analogous to that of *RB* is *adenomatous polyposis coli*, or *APC*.

APC is a tumor suppressor gene that is critically involved in the development of colorectal cancers. Like *RB, APC* was cloned by virtue of its chromosomal location. The first indication of the position of the *APC* locus arose from studies of patient with familial adenomatous polyposis (FAP), a heritable form of colorectal cancer in which predisposed individuals develop a large number of polyps and cancers (see Chapter 1). Upon cytogenetic analysis of normal blood cells,

one FAP patient was identified that had an interstitial deletion within the long arm of chromosome 5. This chromosome had been inherited from an affected parent. As was the case with retinoblastoma, the finding of a constitutional deletion in a susceptible individual provided the first clue as to the disease locus.

A second line of evidence implicating the 5q region in colorectal tumors arose from the study of LOH in sporadic colorectal cancers. These studies found chromosome 5 losses to be a frequent event. Importantly, LOH on chromosome 5 is found in both small adenomas and in large carcinomas, suggesting that inactivation of a tumor suppressor gene on chromosome 5 is an early and perhaps initiating event in the process of tumorigenesis.

A total of four genes were mapped to the common region of loss at chromosome band 5q21–22 by the laboratories of Ray White, Yusuke Nakamura and Bert Vogelstein. Each of these genes was interrogated by DNA sequence analysis. One gene, *APC*, was found to be mutated in sporadic tumors and in the germline of FAP patients. In tumor samples from sporadic and inherited cases alike, LOH had resulted in the complete loss of wild type *APC* alleles.

The cloning of the *APC* gene and the characterization of *APC* mutations in inherited and sporadic forms of colorectal cancer has reinforced and extended many of the basic principles of tumor suppressor genes originally revealed by the cloning of the *RB* gene. The fact that colorectal cancer is much more common than retinoblastoma and occurs in more diverse forms has revealed several important characteristics and types of tumor suppressor gene mutations.

The *APC* mutations that are present in the germline of FAP kindreds and those that occur somatically in sporadic cases are similar in type. Single nucleotide substitutions within the open reading frame cause nonsense codons and splice site mutations, while small insertions or deletions lead to frameshifts. In contrast to *RB* mutations in retinoblastoma, which are most commonly deletions, the majority of mutations that inactivate *APC* result in the truncation of the expressed APC protein.

The clinical features of FAP can vary, depending upon the specific mutant of *APC* that is inherited (see Fig. 3.7). Truncating mutations that occur between codons 463 and 1387 cause retinal lesions called congenital hypertrophy of the retinal pigment epithelium (CHRPE). In contrast, truncating mutations between codons 1403 and 1578 are associated with desmoid tumors and mandibular osteomas, a condition known as Gardner's syndrome, but not with CHRPE. Other specific forms of mutant *APC* are associated with increased or decreased numbers of tumors. Though germline *APC* mutations cause all of these conditions, the different alleles of *APC* are clearly not equivalent. FAP thus illuminates the general principle that the genotype/phenotype relationship involving a tumor suppressor gene and a cancer can be highly specific.

Even patients that carry identical *APC* mutations can have differing disease manifestations. For example, while one individual with a mutation near codon 1500 may develop the extracolonic tumors associated with Gardner's syndrome, another individual with the same mutation may not. Other genetic differences not involving *APC* are likely to play a role in the types of cancer that ultimately arise

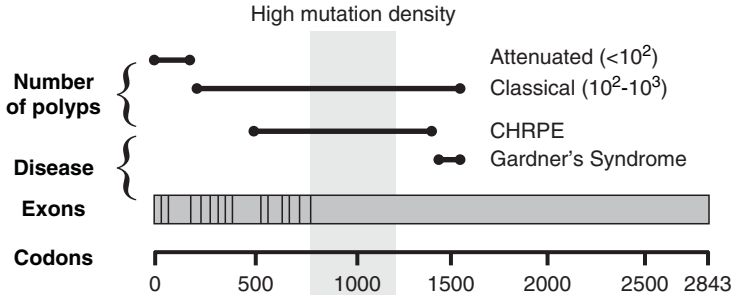

Fig. 3.7 Inherited mutations of the *APC* gene. In many cases, specific APC mutations can be correlated with distinct phenotypes and disease subtypes. The APC gene is composed of 15 exons that in total contain a 8,538 bp open reading frame that encode a full-length protein of 2,843 amino acids. The last exon is unusually large. Approximately two thirds of the total mutations found are clustered in the 5′ region of exon 15. An attenuated form of FAP is associated with mutations within the first 157 codons, whereas more typical pronounced levels of polyposis are associated with mutations occurring between codons 169 and 1600. Mutations within distinct domains cause CHIRPE and Gardner's syndrome

in predisposed individuals. Genetic factors that affect the diseases caused by known tumor suppressor mutations are referred to as *modifiers*.

In the general population, colon polyps are extremely common. About one half of all individuals will develop an adenomatous polyp by the age of 70. Numerous studies have found that heritable factors could account for as many as one half of all colorectal cancers, implying that many of the cases that are now termed sporadic actually have an inherited component. Among patients with multiple adenomatous polyps, approximately 25% carry variant *APC* alleles in their germline, compared with 12% of normal controls. These variants are not truncated *APC* alleles, but are single nucleotide substitutions that may cause subtle variations in APC function. It thus appears to be a possibility that common variants of *APC* carry an increased risk of polyps and cancer. While this hypothesis is compelling, the link between common alleles and this common type of tumor remain to be firmly established.

While the majority of *APC* mutations give rise to truncated and highly dysfunctional proteins, other rare mutations appear to affect cancer predisposition in more subtle ways. These types of mutations do not appear to account for a large proportion of colorectal cancers, but they nonetheless illustrate the remarkable ways that inherited tumor suppressor gene mutations can alter the overall cancer risk:

Mutations causing altered gene expression. An FAP kindred was identified in which no obvious mutation could be detected within the *APC* open reading frame. However, detailed analysis of one affected individual revealed markedly reduced expression of one *APC* allele. Presumably, an unidentified mutation in a regulatory region outside the open reading frame had altered the expression of *APC*. As is typically the case with FAP, LOH was detected in the majority of the tumors removed from this patient. In nearly every instance, the *APC* allele lost was the one

that was expressed at normal levels, that is, the wild type allele. Linkage of a low expressing *APC* allele with polyp formation thus suggests that even partial loss of function of this gene is sufficient to confer cancer predisposition.

Pre-mutations. Distinct germline *APC* mutations cause less penetrant predispositions to colorectal cancer that are distinct from FAP, a highly penetrant syndrome. An *APC* allele that contains a missense mutation in codon 1307, changing the encoded amino acid from an isoleucine (I) to a lysine (K), is found at a relatively high frequency among Ashkenazi Jews. This allele, called 1307K, is present in approximately 6% of individuals in this ethnic group, but is rare in the general population. Among the Ashkenazim, the 1307K allele is overrepresented in patients with colorectal cancer and in individuals with a family history of colorectal cancer. Molecular epidemiology shows that the 1307K allele is associated with a twofold increased risk of developing cancer. The 1307K allele predisposes carriers to cancer by a unique mechanism. Sequence analysis of the codon 1307 region in tumors revealed a high frequency of somatically acquired truncating mutations. These somatic changes were restricted to the 1307K allele; the normal APC allele was not mutated in the tumors analyzed. The surprising conclusion drawn from these studies is that the 1307 germline mutation creates a genomic sequence that is prone to somatic mutations. The 1307 mutation has been accordingly referred to as a pre-mutation, which alone does not significantly alter the encoded protein but instead raises the probability of a subsequent mutation.

P53 Inactivation: A Frequent Event in Tumorigenesis

While *APC* is a very frequent target of mutation during the early development of colorectal cancer, it was not the first tumor suppressor gene discovered to be involved in colorectal tumorigenesis. That distinction belongs to *P53*, the mutant forms of which are highly prevalent cancer genes found about one half of all colorectal cancers and in a large proportion of many other human malignancies. *P53* mutations are associated with some fraction of almost every type of human cancer.

Unlike *RB* and *APC, P53* was cloned only after the discovery and characterization of its encoded protein. The p53 protein was independently discovered by David Lane and Arnold Levine, and their colleagues, in 1979. During the examination of proteins in cells transformed by the SV40 DNA tumor virus (see Chapter 2), both groups detected a 53 kDa protein that was found to be physically associated with an oncogenic, virus-encoded protein called large T antigen. Because the identity and function of this cellular protein were completely unknown, the protein was named for its molecular weight. Specific antibodies raised against p53 were used to screen gene expression libraries, resulting in the isolation of cDNA clones derived from *P53* transcripts. A *P53* cDNA was then used to isolate a genomic DNA clone, which was found to map to the short arm of chromosome 17.

Several attributes of this new gene suggested that it might play an important role in cancer cells. The most compelling finding was that p53 protein levels were elevated, sometimes to a great extent, in a wide range of cultured cancer cells and tumors. Furthermore, forced overexpression of p53 protein in primary cells contributed to changes in that enhanced their tumorgenicity.

The initial discovery of *P53* predated the tumor suppressor gene concept that arose from studies of *RB* and *APC*. At the early stages of *P53* gene characterization, the oncogene hypothesis was well established while evidence favoring a major role for tumor suppressor genes was still very limited. The properties attributed to p53 expression appeared to represent a gain-of-function consistent with a role for *P53* as a proto-oncogene. However, inconsistencies arose in these studies, leading to their reevaluation. It became apparent that *P53* clones isolated by different laboratories had slightly different sequences, which caused differences in the encoded proteins. *P53* DNAs that had been thought to be wild type were actually mutants. These mutated alleles had been inadvertently isolated because *P53* mutant genes are very prevalent in human cancer, a fact that was not yet known at the time.

The confusion over the role of *P53* was resolved with the discovery that the gene mapped to a common region of loss in colorectal cancers, located at 17p13. In a colorectal tumor that had undergone LOH in this region, it was found that the remaining allele of *P53* contained a missense mutation. This single nucleotide substitution was not present in the normal tissue of the same patient and was therefore somatic. These findings perfectly fit Knudson's hypothesis, and thus provided conclusive evidence that *P53* was in fact a tumor suppressor gene.

Subsequent analysis of *P53* genes in large numbers of tumors revealed that loss of *P53* is a frequent event in human cancers. A significant proportion of tumors arising in many different tissues carry somatic mutations that inactivate *P53* (see Fig. 3.8). Among tumors in which *P53* mutations are prevalent are some of the most common forms of cancer. Overall, *P53* is mutated in about half of all human cancers.

In contrast to *RB*, which tends to be inactivated by large deletions, *P53* is typically inactivated by small alterations (see Fig. 3.9). A smaller proportion of mutations inactivate *P53* by truncating the open reading frame, either by a nonsense point mutation or by a small insertion or deletion that causes a frameshift. In some cancer types, specific mutations in *P53* can be correlated with environmental mutagens (see Chapter 1). Ultraviolet radiation, food-borne toxins, and cigarette smoke have all been found to leave highly characteristic mutations in *P53*.

Mutations in *P53* do not occur with equal frequency throughout the coding sequences, but rather typically occur in hot spots that interfere with the functions of the encoded protein. The *P53* gene encodes a 393 amino acid protein. About 20% of all somatic mutations alter one of three codons, 175, 248 or 273. As will be described in detail in Chapter 5, p53 is a transcription factor that binds to specific sequences within the promoters of growth inhibitory genes and activates their transcription. The majority of inactivating mutations occur in exons that encode a large, centrally located, DNA-binding domain that spans codons 100–300. Mutant p53 proteins fail to specifically bind DNA and thus lose their ability to transactivate transcription.

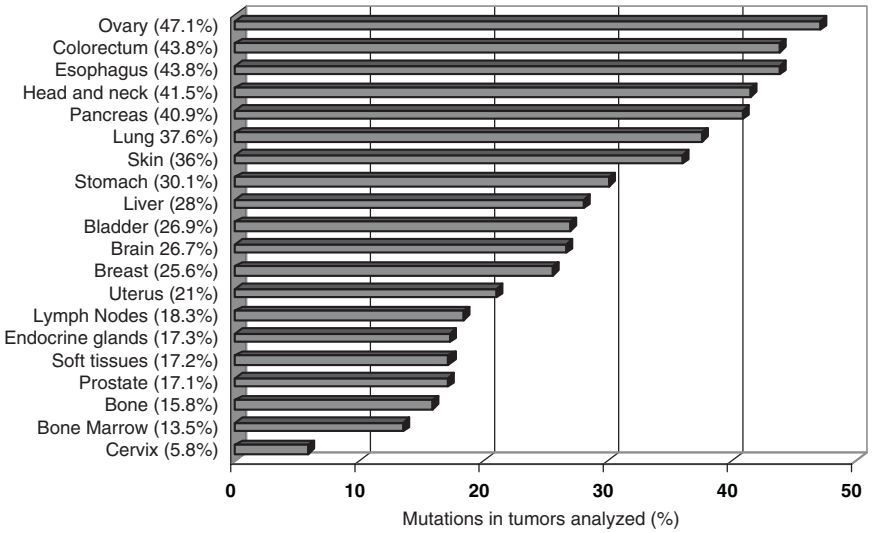

Fig. 3.8 *P53* mutations in human cancers, by site. Mutated *P53* genes are found in many types of cancers. Note that mutations can be difficult to detect and the numbers shown are probably an underestimation of the true frequency. (Data from the IARC TP53 mutation database, R10 release, July 2005.)

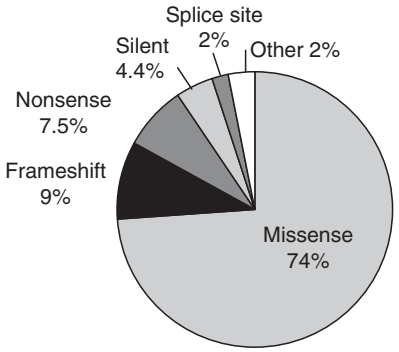

Fig. 3.9 Spectrum of *P53* mutations. The majority of P53 mutations detected in tumors are single nucleotide substitutions that result in missense codons. (Data from the IARC TP53 mutation Database, R10 release, July 2005.)

Functional Inactivation of p53: Tumor Suppressor Genes and Oncogenes Interact

In a significant number of cancers, *P53* is inactivated not by mutation, but by the activation of an antagonistic oncogene. This form of inhibition occurs at the post-translational level and is mediated by protein–protein interactions. There are two highly illustrative oncoproteins that inhibit p53 and contribute to tumorigenesis:

Mdm2. In normal cells, the level of p53 protein is highly regulated. A central mechanism of regulation involves the physical interaction of p53 with a protein encoded by *MDM2*. *MDM2* is a proto-oncogene that was originally found in double minutes in tumorigenic mouse cells (see Chapter 2). The human homolog (sometimes called Hdm2) is an enzyme that covalently modifies proteins by the addition of ubiquitin. The ubiquitination of proteins by ubiquitin ligases like Mdm2 serves to target those proteins for degradation by the proteosome. Thus, the interaction of p53 with Mdm2 leads to p53 degradation, keeping p53 protein levels within a narrow range of intracellular concentration. In several types of cancers, principally soft-tissue sarcomas, the *MDM2* gene is amplified. The increased levels of Mdm2 are oncogenic, causing decreased levels of p53 and resulting in a loss of p53 function. *MDM2* is amplified in roughly one third of sarcomas. The *MDM2* locus is often amplified 50-fold or greater in these cancers.

More subtle alterations in *MDM2* can also affect p53 function. A common allele of *MDM2* contains a SNP that affects expression. A T→G change found at the 309th nucleotide of the first intron, within the gene promoter, increases the binding of a transcriptional activator and results in higher levels of *MDM2*-encoded RNA and protein. The increased level of Mdm2 causes a corresponding attenuation of p53 function. Known as the SNP309 allele, this *MDM2* variant is very common; approximately 40% of individuals in the general population are heterozygous and 12% are homozygous for SNP309. Studies of patients with sporadic soft tissue sarcomas revealed that those homozygous for SNP309 had an age of disease onset that was at least 9 years earlier, and that that young sarcoma patients had a higher frequency of SNP309. These findings suggest that attenuation of p53 function can accelerate tumor formation. It might be expected that SNP309 confers an increased cancer risk, but such a link has not been firmly established.

Human papillomavirus oncoprotein E6 (HPV E6). A second oncogene that affects p53 is not a cellular gene, but rather is a viral introduced upon infection by the human papillomaviruses (HPV). While the vast majority of cancers arise solely as a result of germline and/or somatically acquired mutations, an atypical exception is cancer of the uterine cervix. In the majority of cervical cancers, p53 is functionally inactivated by the inhibitory binding of an HPV-encoded protein known as E6. The role of HPV in cervical tumorigenesis will be described in Chapter 6.

Germline Inheritance of Mutant *P53*: Li–Fraumeni Syndrome

Like *RB* and *APC, P53* mutations are involved in both sporadic and inherited cancers. *P53* mutant alleles are found in the germline of individuals with an inherited susceptibility to cancer known as Li Fraumeni syndrome (LFS). LFS is an autosomal dominant disorder mainly characterized by the early onset of bone or soft tissue sarcomas.

LFS was first recognized as a clinical entity by Frederick Li and Joseph Fraumeni, in 1969. Five kindreds were identified in which childhood sarcomas affected siblings

or cousins. Soon after the initial report by Li and Fraumeni, Lynch and colleagues similarly described pedigrees with unusual clusters of diverse cancers, including sarcomas, breast, lung, laryngeal and brain cancer, and leukemias.

Cancer predisposition syndromes that involve diverse types of cancer can be difficult to classify. Such was the case with LFS, which for a time was known by varying terminology. The precise clinical criteria for LFS took years to firmly establish largely because of inherent biases in the ways that clinical syndromes come to the attention of epidemiologists. Known as *ascertainment biases*, factors that complicate the classification of cancer syndromes include the preferential attention paid to kindreds that are most dramatically affected, the clustering of common cancers in families purely by chance, and uncertainty regarding the prevalence of the syndrome and the penetrance of the underlying genetic defect. In the case of LFS, these factors were eventually mitigated by the establishment of rigorously defined diagnostic criteria and, ultimately, by the identification of the inherited cancer genes.

As was observed in the inherited form of retinoblastoma, many patients with LFS were found to develop multiple primary tumors. Additionally, cancer was often found to strike at several times throughout life in LFS patients. In many cases cancers occurred years apart. The later-onset cancers were often causally related to previous rounds of cancer therapy. While the epidemiological data were strongly suggestive of an inherited predisposition to cancer, the molecular basis of this predisposition was unknown.

The discovery of *P53* mutations in many different types of cancer, including those that commonly affect LFS patients, prompted the examination of the *P53* alleles in LFS kindreds. Mutations in *P53*, primarily single nucleotide substitutions, were found in the germline of all of the affected individuals from the initial five kindreds tested. Analysis of tumors from these patients confirmed that the normal copy had been lost during the process of tumorigenesis; tumors were homozygous for the mutant *P53* allele. Not all individuals that carried the mutant *P53* allele had been affected by cancer at the time that they were tested. These individuals were at high risk for developing cancer in the future. The identification of *P53* mutations as the genetic defect that underlies Li Fraumeni syndrome was a historical landmark: for the first time, cancer predisposition could be predicted by genotype analysis.

Subsequent analysis of many LFS kindreds has revealed that the spectrum of *P53* mutations found in LFS is similar to that found in sporadic cancers: about three quarters are missense mutations (see Fig. 3.10). Similarly, the mutations of *P53* associated with LFS typically affect the central, DNA-binding domain of the encoded protein. There is considerable overlap: codons 248 and 273 are most commonly mutated in both sporadic cancers and in LFS. However, some of the codons mutated in LFS, such as codon 337 which is mutated in about 10% of disease kindreds, are rarely mutated in sporadic cancers.

The patterns of cancer that occur in LFS patients are partially dependent on the precise *P53* mutation inherited. Mutations within the exons that encode the DNA-binding domain of p53 (see Chapter 5) are associated with a higher prevalence of

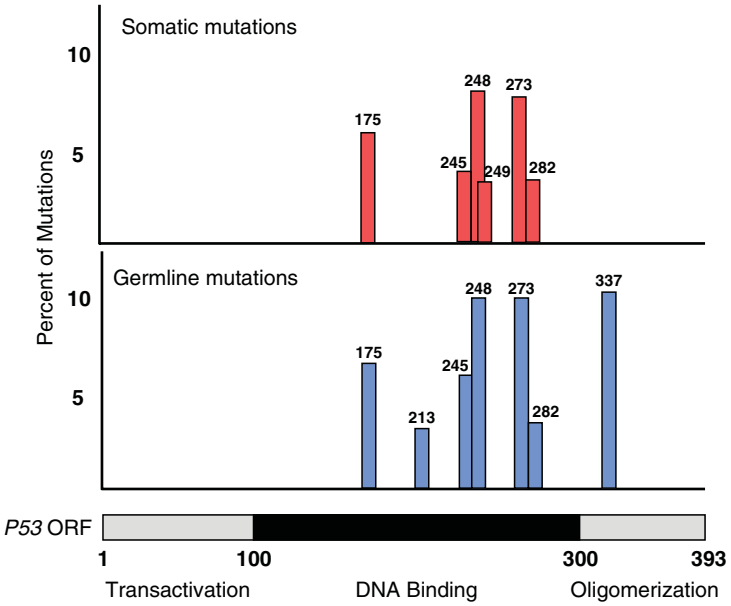

Fig. 3.10 Somatic and inherited mutations of *P53*. The distribution of *P53* mutations in Li Fraumeni kindreds and in sporadic tumors is similar. Most mutations affect the central region that encodes a sequence-specific binding domain critical for protein function. A highly acidic domain that interacts with other transcription factors is rarely targeted by mutation. A c-terminal domain is important for the organization of p53 molecules into active, oligomeric complexes. A relatively common germline mutation in this coding region is rarely found in sporadic tumors. Note that only the most common mutations (>3%) are shown. More rare mutations are generally clustered in the DNA-binding domain

brain tumors and an earlier onset of breast cancers, whereas mutations outside the DNA-binding domain are associated with a higher incidence of adrenal cancers.

The tumors that occur in affected Li Fraumeni individuals are similar in type, but not identical in proportion, to the sporadic cancers that contain *P53* mutations (see Fig. 3.11). Breast cancer eventually occurs in over a quarter of LFS patients; a similar proportion of spontaneous breast cancers have somatic mutations in *P53*. However, while *P53* is mutated in almost one half of all spontaneous colorectal cancers, this type of tumor is infrequently seen in LFS families. The initiating event of the colorectal cancers in Li Fraumeni cases is loss of *APC* function, just as in the general population.

Mutations of *P53* have not been found in all families that fit the clinical criteria for LFS. This lack of complete concordance between the *P53* mutant genotype has several possible explanations:

False negatives. The techniques of analyzing patient-derived DNA samples have been in a state of constant development and refinement. Older, less reliable DNA sequencing techniques were less sensitive, more laborious and more expensive. Because of these limitations, it was routine to analyze only limited regions of the *P53* locus. In only a subset of LFS kindreds has the *P53* locus been completely analyzed

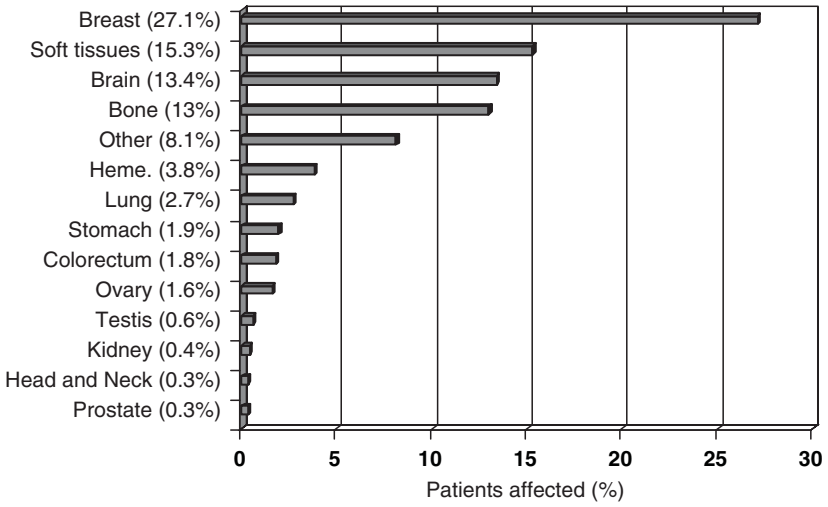

Fig. 3.11 Tumors associated with germline *P53* mutations. Individuals with Li Fraumeni syndrome are predisposed to diverse cancers. (Data from the IARC TP53 mutation database, R10 release, July 2005.)

with highly sensitive technology. The lack of sensitivity and the exclusion of large regions of genomic DNA in most sequencing efforts to date have contributed to an underestimation of germline *P53* mutations in LFS kindreds. Obsolete technology causes mutations to be missed.

Phenotypic variability. The aforementioned ascertainment biases complicate patient categorization, and can lead to errors in identifying true LFS families. The establishment of rigorous diagnostic criteria has been important in minimizing this problem. Careful diagnosis has resulted in the identification of a separate group of families in which the classic phenotype of LFS is incompletely expressed. This group has accordingly been termed Li Fraumeni-like syndrome (LFLS). *P53* mutations have been found in some of these kindreds as well, suggesting that the penetrance of *P53* mutations can vary in different genetic backgrounds.

Other mutations. p53, like all cancer gene products, functions as part of a multiprotein pathway (see Chapter 5). Mutations in other genes that contribute to the p53 pathway could theoretically phenocopy some of the effects of *P53* mutation. Several genes have been proposed as putative tumor suppressor genes that might account for some LFS kindreds, but the mutational data has not been conclusive.

The variability of LFS, even in rigorously defined cases with validated *P53* mutations, suggests that there are other genetic factors, known as modifiers, which can affect the *P53* mutation-carrier phenotype. One modifier of *P53* is the common SNP309 allele of *MDM2*. SNP309 occurs at a frequency in LFS kindreds that is similar to that in the general population. But LFS patients that also carry the SNP309 allele develop cancers at a significantly earlier age than FLS patients without SNP309. As described above, SNP309 increases the

Mdm2 level and thereby reduces the level of p53. It is important to note that LOH at the *P53* locus is the rate limiting step of for tumor formation in LFS, regardless of *MDM2* genotype. The SNP309 results thus imply that a reduction in p53 protein levels lead to an increased frequency of *P53* allelic loss. How these two disparate events might be linked is an important but unresolved question.

While somatic mutations that affect *P53* are very common in cancers, the frequency of germline *P53* mutant alleles within the human population is very low. LFS is accordingly rare. Overall, *P53* mutations contribute to many more sporadic than inherited cancers. Nonetheless, LFS serves to illustrate a central principle of cancer genetics. In familial retinoblastoma, FAP and Li Fraumeni syndrome, inheritance of a mutated tumor suppressor gene is followed by loss of the single wild type allele in the cells of the developing tumor. Thus, the pattern of inactivating mutations in sporadic and inherited cancers conforms to the predictions of Knudson's hypothesis.

Cancer Predisposition: Allelic Penetrance, Relative Risk and Odds Ratios

That tumor suppressor genes are mutated in sporadic cancers and are also inherited in the germline of cancer-prone families is incontrovertible evidence of their central importance in tumorigenesis. Numerous well-defined cancer syndromes are understood at the genetic level (see Table 3.1). Collectively, these syndromes account for a small but significant proportion of all human cancers. Understanding the genetic basis of a cancer syndrome allows carriers within known kindreds to be firmly identified. Cancer syndromes also illuminate the etiology of sporadic cancers. With few exceptions, the sporadic cancers in which a somatic mutated gene is predominantly found mirror those that are characteristic of the inherited syndrome associated with that gene. The relationship between inherited and sporadic

Table 3.1 Tumor suppressor gene mutations in inherited cancer syndromes and in sporadic cancers. The predominant forms of inherited cancers are indicated in bold type

Gene	Cancer syndrome	Penetrance*	Inherited cancers	Sporadic cancers
RB	Familial retinoblastoma	>95%	**Retinoblastoma** **Osteosarcoma** **Neuroblastoma** **Melanoma**	Retinoblastoma Endometrial Bladder Osteosarcoma, Lung
APC	Familial adenomatous polyposis, Gardner's syndrome	>95%	**Colorectal** **Osteosarcoma** **Small intestinal** **Gastric**	Colorectal Gastric Small intestinal Adrenal gland Pancreatic

(continued)

Table 3.1 (continued)

Gene	Cancer syndrome	Penetrance*	Inherited cancers	Sporadic cancers
P53 (TP53)	Li Fraumeni syndrome	>95% females; ~75% males	Breast Sarcoma Brain tumors Osteosarcoma	Ovarian Colorectal Ca Esophageal Head and neck Pancreatic Lung Skin Breast Endometrial Lymphoma
PTEN (MMAC1, TEP1)	Cowden syndrome, Bannayan-Riley-Ruvalcaba syndrome	>95%	Breast Thyroid Endometrial Brain	Endometrial Brain Prostate Lung Breast Bladder Ovarian Lymphoma
BRCA1	Familial breast and Ovarian Ca	~80% breast; ~40% ovarian	Breast Ovarian	Ovarian Breast (rare)
BRCA2	Familial breast Ca	~80% breast; ~20% ovarian	Breast (inc. males) Ovarian Pancreatic	Breast (rare) Colorectal (rare)
NF1	Neurofibromatosis Type 1	>95%	Brain Neural tumors	Melanoma Neuroblastoma
NF2	Neurofibromatosis Type 2	>95%	Neural tumors	Brain tumors
VHL	von Hippel-Lindau syndrome	>60%	Kidney Hemangioblastoma	Kidney Hemangioblastoma
MEN1	Multiple endocrine neoplasia	~90%	Pancreatic islet Cell	Pituitary Adenomas Parathyroid
SMAD4 (DPC4)	Familial juvenile polyposis syndrome	~20%	Colorectal Gastric Small intestinal	Pancreatic Colorectal
CDKN2A (P16, INK4, MTS1)	Familial melanoma	~70%	Melanoma Pancreatic Breast	Melanoma Pancreatic Esophageal Lung Head and neck Leukemia Bladder
MSH2, MLH1, MSH6, PMS2	Hereditary nonpolyposis colorectal cancer, Turcot syndrome	~80% colorectal; ~70% endometrial	Colorectal Endometrial Ovarian Small intestinal Bladder Brain Biliary tract	Colorectal Gastric Endometrial Bladder

*Lifetime risk for developing the predominant form of cancer.

forms of cancer thus provides important clues into the molecular pathogenesis of the specific cancers involved.

Importantly, genetically defined cancer syndromes illustrate how the risk of cancer is inherited. The inheritance of a cancer gene allows cancer cell precursors to bypass a step on the genetic path to tumor formation (see Chapter 1). For rate-limiting steps, the extent of cancer predisposition can be striking. Inheritance of an inactivated allele of *APC* confers a virtual guarantee that, without prophylactic therapy, that individual carrier will develop malignant colorectal cancer at a young age. Other germline alleles confer risks that are less obvious.

Three parameters describe the consequence of inheriting a germline tumor gene mutation: the *penetrance* of the disease phenotype, the *relative risk* of developing cancer and the *odds ratio*. These figures are related to one another, but dependent on distinct variables:

Penetrance. The penetrance of a mutant tumor suppressor gene and the absolute risk of cancer conferred by that mutation are one and the same. For example, inheritance of a gene that has a penetrance of 50% imparts an absolute risk that is also 50%. One half of the carriers of that allele will develop cancer. In cases of incomplete penetrance, additional genetic and environmental factors that are difficult to quantify will play an important role in determining which individuals will develop disease. When penetrance is near-complete, as is the case with familial retinoblastoma, other genes and environmental factors are less relevant to the absolute risk. It is important to note that different alleles of the same tumor suppressor gene can be differentially penetrant, as is the case with the breast cancer susceptibility genes described below.

Relative risk. All human beings are at risk of cancer. In kindreds with germline tumor suppressor gene mutations, that risk is elevated. For a given cancer, the relative risk (also known as the risk ratio) compares the probability of cancer in two groups and is defined as:

$$\text{Relative risk} = \frac{\text{Absolute risk of cancer in carriers (\%)}}{\text{Absolute risk of cancer in the general population (\%)}}$$

Odds ratio. Another comparison of risk between two cohorts is the odds ratio. Most often applied to case-control studies in which the outcome (i.e. cancer) is a rare occurrence, the odds ratio compares the relative odds of cancer between two groups. Applied to the analysis of carriers of a specific allele:

$$\text{Odds ratio} = \frac{\text{Odds against developing cancer in the general population}}{\text{Odds for developing cancer in carriers}}$$

As an example, consider a hypothetical cancer-causing allele that has a penetrance of 10%. If the incidence of the same cancer in the general population is 1%, the relative risk is 0.10/0.01 = 10. In contrast, the odds ratio would be (99 to 1)/(1 in 10) = 99/0.1 = 990. In general, the relative risk yields a more intuitive result than an odds ratio, but can lead to misleading results if applied to studies in which only the outcome is measured.

Breast carcinoma is a common type of cancer that can be caused by incompletely penetrant, mutant tumor suppressor genes. The relative risk and odds ratio associated with each inherited allele is highly meaningful in the interpretation of carrier status. Individuals heterozygous for a tumor suppressor gene with a known penetrance can be counseled as to their risk of cancer and monitored accordingly. In cases where inherited mutations have a high penetrance, carriers may opt for prophylactic therapy. Unlike FAP which features polyposis, inherited breast cancer syndrome does not feature a readily detectable heterozygous phenotype. Carriers of breast cancer susceptibility alleles must therefore be identified on the basis of their genotype. The use of genetic information for population screening and risk assessment will be discussed in Chapter 7.

Aside from the known syndromes listed in Table 3.1 and described in this chapter, there are many familial clusters of cancer that are less well understood. While highly penetrant genes that cause readily discernible forms of hereditary cancer are most straightforward to classify, tumor suppressor genes with incomplete penetrance that contribute to common forms of cancer can be much more difficult to detect. In the near future, the application of large-scale genotype analysis to cancer-prone families and sporadic tumors promises to have a large impact on both tumor suppressor gene discovery and risk analysis.

Breast Cancer Susceptibility: *BRCA1* and *BRCA2*

Breast cancer is among the most common malignancies. Like most cancers, the majority of cases are sporadic. However, epidemiologic evidence has long supported an inherited component for a small proportion of breast tumors (approximately 5%). The identification of *RB* and *APC* created a paradigm for relating both sporadic and inherited forms of a cancer to a single tumor suppressor gene. It was anticipated that analysis of the small fraction of inherited breast cancers might be similarly informative.

From a genetic perspective, breast cancer poses a major challenge. There is no singular, well defined syndrome featuring near-complete penetrance, as is the case with FAP and familial retinoblastoma. Overall, the clinical presentation of inherited and sporadic breast cancers is largely similar. Key features of inherited breast cancers are bilateral tumors and the onset of disease prior to menopause; these features can often be overlooked. Furthermore, because breast cancer is so common, it can be difficult to firmly identify kindreds that carry a predisposition. While a familial cluster of retinoblastoma is a reliable indicator of inherited susceptibility, multiple cases of breast cancer can occur in a single family solely by chance. Further complicating genetic analysis, sporadic cancers can and do occur in kindreds that carry a predisposing mutation in the germline; such cancers will occur in carriers and non-carriers alike. Unraveling the genetic basis of breast cancer is an ongoing process that has benefited from an approach that combines careful epidemiology with molecular genetic analysis.

P53 was the first breast cancer gene to be described. *P53* mutations are present in a significant proportion – but not the majority – of sporadic breast cancers (see Fig. 3.8). Breast cancer is a primary phenotype of Li Fraumeni syndrome, but because breast cancer is common and Li Fraumeni syndrome is relatively rare, Li Fraumeni cases do not account for a significant proportion of the total cases.

In pursuit of more common breast cancer genes, investigators sought chromosomal markers that were genetically linked to early onset cases within familial clusters. Focusing on a large group of families cumulatively composed of thousands of cases of early-onset breast cancer, Mary-Claire King and her coworkers established linkage with a region on the long arm of chromosome 17 in 1990. Three years later, Mark Skolnick and colleagues identified a gene in this region, termed *BRCA1*, which was mutationally truncated in the germline of several kindreds. *BRCA1* mutations were subsequently found in a major proportion of previously identified families with high incidence of inherited breast as well as ovarian cancers. By examining the families that did not carry mutant *BRCA1*, a large consortium of investigators found linkage to a second breast cancer susceptibility locus on chromosome 13. The *BRCA2* gene was cloned in 1995. In total, mutations in either *BRCA1* or *BRCA2* are thought to contribute to more than one half of inherited breast cancers.

The two breast cancer susceptibility genes are structurally unrelated. *BRCA1* is composed of 24 exons that encode a 1863 amino acid protein. Almost one half of the germline mutations are single base substitutions that include missense, nonsense and splice site mutations. The remaining *BRCA1* mutations are predominantly small deletions and insertions. Truncation of the open reading frame is a common consequence of *BRCA1* mutation. Mutations have been detected throughout the *BRCA1* coding sequences. *BRCA2* is a 27-exon gene that encodes a 3418 amino acid protein. As is the case with *BRCA1*, the mutations in *BRCA2* are often truncating mutations caused by single nucleotide substitutions and small insertions and deletions. Among *BRCA2* mutations, those in the central region of the gene appear to confer a higher risk of ovarian cancer. This region has been termed the ovarian cancer cluster region.

Both *BRCA1* and *BRCA2* are high-penetrance, dominant cancer genes. The exact penetrance of *BRCA1* and *BRCA2* mutations, and therefore the risk of cancer associated with such mutations, has been difficult to ascertain for several reasons. Different mutations appear to confer somewhat distinct risks. Another complicating factor is that the average age of incidence can vary significantly among families with the same mutation. The patterns of cancer can also vary. Some families have increased incidence of breast cancer only, while other families with the same mutation can present with breast and ovarian cancers. The reasons for this high degree of variability in risk are unknown, but probably to relate to both modifying genes and to components of lifestyle and environment. The lifetime risks of breast and ovarian cancer associated with *BRCA1* and *BRCA2* mutations are shown in Table 3.2. Carriers of either *BRCA1* or *BRCA2* have a nearly sevenfold higher risk for developing breast cancer during their lifetimes. This relative risk may seem somewhat low for a high-penetrance gene, but this is a direct result of the high incidence of sporadic breast cancers. Indeed, the relative risk conferred by a *BRCA1* or *BRCA2* mutation

Table 3.2 Lifetime risks for developing cancer associated with *BRCA1* and *BRCA2* mutations. The penetrance of *BRCA1* and *BRCA2* mutations has been found to be highly variable. Figures shown are representative but highly approximate

		Mutant *BRCA1* carrier		Mutant *BRCA2* carrier	
Cancer	General population	Risk	Relative risk*	Risk	Relative risk*
Breast	12%	80%	6.7	80%	6.7
Ovarian	1.8%	40%	22	20%	11
Male breast	0.1%	3%	30	6%	60

*Defined as the fold-increase in the overall risk attributable to the mutated gene.

for the onset of breast cancer before age 40, which rarely occurs sporadically, is roughly 150. Approximately 60–80% of female carriers of *BRCA1* or *BRCA2* mutations develop breast cancer during their lifetimes.

Male carriers of *BRCA2* mutations are also at an increased risk of breast cancer and possibly prostate cancer. Overall, male breast cancers are rare, with a prevalence that is about 1% of all breast cancers. About 30% of affected individuals have male or female relatives with breast cancer, suggesting that male breast cancer has a significant heritable component.

While *BRCA1* and *BRCA2* germline mutations are diverse, several mutations have been found to be present in different families. These recurrent mutations are typically restricted to specific ethnic groups and are thought to reflect what is known in genetics as a *founder effect* – a recurring trait in a growing population that originates from a small group of common ancestors. Founder mutations in *BRCA1* and *BRCA2* have been found in Jewish, Icelandic and Polish populations. Three different founder mutations have been found in individuals of Ashkenazi Jewish ancestry, and are present in about 2% of that population.

Mutations in *BRCA1* or in *BRCA2* occur at a frequency of approximately 1 in 250 women, suggesting that 250,000 women in the USA are carriers. The relatively low frequency of mutations in the general population and the clustering of founder mutations in defined ethnic groups has significant implications for the use of genetic screens to identify individuals at risk for cancer. These issues will be discussed further in Chapter 7.

BRCA1 and *BRCA2* are not widely mutated in sporadic breast cancers or other cancers. This was a surprise and in some respects a disappointment. *RB, APC* and *P53* mutations are centrally involved in both inherited and sporadic forms of cancer, and it was widely assumed that solving the basis of hereditary breast cancer would similarly inform an understanding of the much more common sporadic cancers as well. This has not turned out to be the case. While the cloning of the *BRCA1* and *BRCA2* genes was a technological triumph, the results of that successful effort have not been directly applicable to the pathogenesis of most breast tumors. Nonetheless, the functional analyses of *BRCA1* and *BRCA2* have provided useful insights. The proteins encoded by *BRCA1* and *BRCA2* play important roles

in the repair of damaged DNA, suggesting that their tumor suppressor function is based upon the suppression of spontaneous mutations. As epidemiological studies have pointed to strong links between breast cancer incidence and environmental mutagens, the involvement of *BRCA1* and *BRCA2* in hereditary forms of the disease suggest a compelling relationship between mutagenesis, DNA repair and breast tumorigenesis. The nature of DNA repair pathways and their role in breast cancer will be discussed in detail in Chapter 5.

Genetic Losses on Chromosome 9: *CDKN2A*

Another frequent site of genetic loss in human cancers is the short arm of chromosome 9. Cytogenetic abnormalities affecting region 9p21 are found in numerous tumors, including melanomas, leukemias and brain and lung cancers. In 1994, a group led by Dennis Carson examined the patterns of loss of two known genes within this region that were variably deleted in cancer cell lines, and determined that a tumor suppressor resided between them. Mapping and sequencing of this region revealed a gene that was consistently deleted in sporadic cancers. An independent group, led by Mark Skolnick, isolated the same gene by mapping homozygous deletions in melanoma cell lines. In melanoma cell lines that had lost only a single allele, the remaining allele was frequently found to harbor a nonsense, missense or frameshift mutation. These pieces of evidence were a strong indication that a new tumor suppressor gene had been found.

It was immediately apparent that the gene on 9p21 encoded a protein that was already known to play a central role in the regulation of cell growth. A year before the positional cloning of the 9p21 tumor suppressor locus, David Beach and his coworkers had discovered and characterized a 16 kDa protein, designated p16, that binds to cyclin-dependent kinase 4 (Cdk4), an enzyme that promotes the progression of the cell cycle. The binding of p16 to Cdk4 inhibits this activity. Sequencing the open reading from of the 9p21 tumor suppressor gene quickly revealed that p16 is the encoded protein. The p16 proteins encoded by the tumor derived mutants failed to inhibit Cdk4 and thus failed to block cell cycle progression. Interestingly, an important downstream substrate of Cdk4 is RB, the product of the tumor suppressor gene inactivated in familial retinoblastoma. The compelling functional link between p16 and RB suggested that inactivation of their corresponding tumor suppressor genes might have similar cellular effects. The relationship between p16, RB and the progression of the cell cycle will be described in detail in Chapter 5.

The tumor suppressor gene on 9p21 was designated *CDKN2A*, to reflect the role of the encoded protein as a specific inhibitor of Cdk4 and as a member of a family of genes that are cyclin dependent kinase inhibitors. The protein encoded by *CDKN2A* is still referred to as p16.

CDKN2A is mutated in a wide range of sporadic tumors. Melanomas are the form of cancer most commonly associated with *CDKN2A* loss. About 20% of sporadic melanomas homozygously inactivate *CDKN2A*. A common form of

CDKN2A inactivation is deletion, but significant proportions of missense, nonsense and insertion mutations also occur. *CDKN2A* mutations are commonly seen in pancreatic, esophageal, lung, head and neck, and bladder cancers and in some leukemias.

Exposure to UV is an important environmental risk factor for melanoma development. A significant number of *CDKN2A* alterations are single base substitutions of the C→T and CC→TT type, which are known UV signature mutations (see Chapter 1). However, the relationship between UV exposure and melanoma risk is complex. As will be described in Chapter 6, UV signature mutations are not consistently observed in other genes that contribute to melanoma tumorigenesis.

Approximately 10% of all melanoma cases are familial, defined as those cases that occur in either two first-degree relatives or in three family members in total, irrespective of the degree of relationship. Linkage to the *CDKN2A* locus has been demonstrated in approximately 50% of melanoma-prone families, though defined *CDKN2A* mutations have been found in only about 20%. The reason for this discrepancy is likely to lie in the technical challenges inherent in detecting large deletions. There are several mutations that appear multiple times in defined subpopulations, and are thus likely to represent founder mutations. Overall, carriers of germline *CDKN2A* mutations have a 75-fold increased risk of developing melanoma, as compared to the general population (relative risk = 75). In addition, *CDKN2A* mutation carriers are also at a significantly higher risk of developing pancreatic cancer, with a relative risk of approximately 22. An increased risk of pancreatic cancer is not apparent in melanoma-prone kindreds that do not have a mutation in *CDKN2A*.

Within melanoma-prone kindreds it is common that melanomas coexist with benign skin lesions known as *atypical nevi*, or *atypical moles*. Prior to the cloning of *CDKN2A*, it appeared that susceptibility to both melanomas and atypical nevi might have a common underlying genetic cause. Unexpectedly, it appears that atypical nevi do not always cosegregate with mutant *CDKN2A*, suggesting that additional genetic factors are also relevant to the development of these lesions. Clearly there are other loci that play important roles in the development of both melanomas and atypical nevi.

Within cancer-prone kindreds, affected individuals are almost always heterozygous for germline mutant tumor suppressor genes. As we have repeatedly seen, the single wild type allele is lost during tumorigenesis, which upon analysis is seen as LOH. Remarkably, two individuals with biallelically mutated *CDKN2A* alleles have been identified. Both were homozygous for a known founder mutation in *CDKN2A* that was present in each of their parents. Every cell in the bodies of these two individuals had thus already sustained two 'hits' of *CDKN2A*. The homozygous patients were cancer-prone but otherwise healthy, indicating that expression of p16 protein is not essential for cellular viability or normal development. One of these patients developed two primary melanomas by the age of 15, while the other was melanoma-free until she died at the age of 55 from an adenocarcinoma. The dramatically different onset of cancer in these two homozygous individuals clearly illustrates the variable penetrance conferred by even a complete absence of *CDKN2A* function.

The typical penetrance of mutant *CDKN2A* in well-defined melanoma-prone kindreds has been somewhat difficult to establish, but appears to be roughly 70%. As shown by the previous example, the same germline mutations can be variably penetrant in different individuals. Additionally, as was found to be the case with *BRCA1* and *BRCA2* in breast cancers, it has become apparent that some germline mutations of *CDKN2A* inherently vary in their average penetrance.

Multiple primary tumors, in any cancer type, are a hallmark of an underlying predisposition. Melanomas, visible on the surface of the skin, can be diagnosed with relative ease compared with internal tumors, and patients with multiple tumors are readily apparent. It had long been noted that a subset of melanoma patients develop multiple lesions with no known family history of melanoma. These cases were classified as sporadic, but the multifocal nature of their primary lesions suggested the germline presence of a low penetrance tumor suppressor gene mutation. Analysis of *CDK2NA* revealed a significant proportion of mutations in such patients. In several cases, close investigation of other family members that carried these mutations revealed evidence of previously obscure family history of the disease. Thus, genetic analysis of *CDKN2A* was used to detect familial patterns of disease that were not previously clinically apparent. These results would suggest that carriers of these low-penetrance alleles have an increased risk of disease and would therefore benefit from close surveillance.

In summary, the analysis of *CDKN2A* mutations has been highly revealing. *CDKN2A* is widely mutated in many types of sporadic cancer. This alone is a good indication that loss of *CDKN2A* is an important contributor to tumorigenesis. While germline mutation of *CDKN2A* is clearly a predisposing factor for cancer development, the presence of a mutant allele is clearly insufficient to guarantee that a cancer will eventually develop.

Complexity at *CDKN2A:* Neighboring and Overlapping Genes

There are two idiosyncrasies that have complicated the analysis of the *CDKN2A* locus and its role in cancer. The first relates to its neighborhood. The second is the highly unusual structure of the locus and the transcripts that are expressed from it. These features of *CDKN2A* illustrate some of the challenges that arise when examining genetic losses and mutations:

CDKN2B encodes a distinct cyclin dependent kinase inhibitor. The *CDKN2A* locus is located immediately adjacent to another gene, *CDKN2B*, which also encodes a protein that inhibits the activity of cyclin-dependent kinases. While the proximity of these two genes may seem to be a highly improbable coincidence, there are in fact numerous examples of genes with related function being closely linked. A widely known example is the cluster of genes that determine histocompatibility on the short arm of chromosome 6. The evolutionary basis for this type of clustering remains incompletely understood, but is likely to involve gene duplication events.

Allelic losses affecting the 9p21 region are commonly observed in many cancers. There are identified kindreds that exhibit 9p21 loss, but the tumors from affected individuals have no detectable *CDKN2A* mutation. One possibility is that a neighboring locus is an alternative target of the first 'hit'. Is *CDKN2B* also a tumor suppressor gene? The indirect evidence is compelling. The *CDKN2B* locus expresses a 15 kDa protein with considerable similarity to p16. Both are found to function in cancer-related pathways that inhibit the progression of the cell cycle. Many of the larger genomic deletions that inactivate *CDKN2A* also affect *CDKN2B*. In one large-scale analysis of sporadic cancers, many deletions were found that affected both genes and several affected *CDKN2A* but left *CDKN2B* intact. Notably, there were no mutations that deleted *CDKN2B* and left *CDKN2A* intact. While inactivating point mutations were found in *CDKN2*A, none were detected in *CDKN2B*. Critically, no germline mutations within *CDKN2B* have been reported. Lacking these types of direct evidence, it is not possible to definitively characterize *CDKN2B* as a tumor suppressor gene.

The CDKN2A alternative reading frame. Another interesting and potentially important feature of the *CDKN2A* locus was reported in 1995, after several groups observed that a second transcript is encoded by *CDKN2A* (see Fig. 3.12). A previously unrecognized exon, designated exon 1β, was found to reside upstream of the first coding exon that encodes p16, exon 1α. Exon 1β is spliced to the same downstream exons that encode p16, but define an alternative reading frame. This unusual transcript encodes a 132 amino acid, 14 kDa protein designated p14(ARF).

The two distinct proteins encoded by *CDKN2A* are not merely splice variants. Because they are encoded by two different reading frames, the primary structures of p16 and p14(ARF) are unrelated. Furthermore, the expression of p16 and p14(ARF) are controlled by separate promoters.

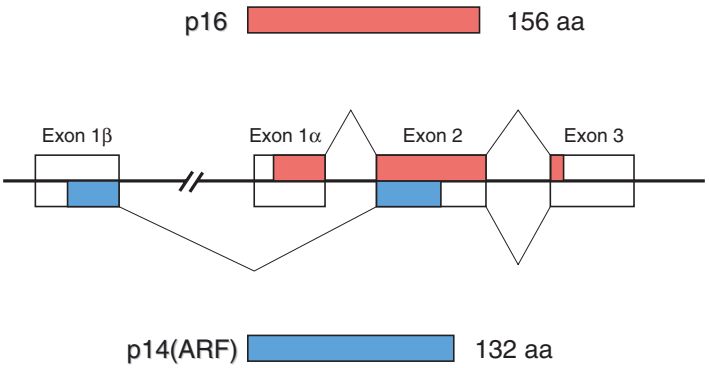

Fig. 3.12 One gene, two proteins. The *CDKN2A* locus is unique in that it encodes two distinct proteins, p16 and p14ARF. The two transcripts originate from two different first exons and use different reading frames within a common exon 2. For these reasons, the proteins are not homologous

Functional analysis of p14(ARF) has shown that it can play a role in the regulation of the cell cycle, and this role is distinct from that of p16. The p14(ARF) protein can bind Mdm2 protein and thereby regulate the levels of the p53 tumor suppressor. Thus, p14(ARF) provides a compelling functional link between two commonly mutated tumor suppressors, p16 and p53.

Is loss of p14(ARF) an important step in tumorigenesis? The overlapping nature of these two genes forced a reevaluation of the mutation data. Most of the point mutations and deletions that affect the p16-encoding exons also affect the p14(ARF) exons. Exon 1β has been found to be selectively deleted in several melanoma cell lines, leaving the p16 coding exons intact. This would be highly suggestive of a role for p14(ARF) in tumor suppression. However, there is some evidence that deletions of 9p21 actually occur during cell culture; the exon 1β deletions observed could therefore represent an artifact. Point mutations within exon 1β have not been detected in tumors, nor have such mutations been found in the germlines of cancer-prone kindreds. In contrast, germline and somatic mutations of exon 1α, specific to the p16 coding region, have been recurrently observed.

Despite the lack of conclusive evidence that loss of p14(ARF) function is critical to human cancer, the unusual structure of *CDKN2A* is interesting in part because it is unprecedented. The complexity of this locus is striking: two unrelated proteins are expressed, via two distinct and independently regulated promoters, from an overlapping exon but in two different reading frames. From an evolutionary perspective, it is difficult to guess how such a locus might have arisen. The p16 protein is highly conserved, as are many cancer genes. The p14(ARF) open reading frame, in contrast, is not more highly conserved between mammals than arbitrary open reading frames, and thus there is little evidence for selective pressure on p14(ARF).

Genetic Losses on Chromosome 10: *PTEN*

The loss of tumor suppressor loci represents an important quantitative difference between cancer cells and their normal precursors. Linkage analysis can best identify these relatively small differences against a background of ‘sameness’. For this reason, positional cloning approaches generally required a high frequency of mutation in a clearly defined type of cancer. Mutations in *APC, RB* and *BRCA1* and *BRCA*2 are highly specific to colorectal cancers, retinoblastomas and breast cancers, respectively. This tumor-specificity, combined with extensive epidemiological data and the identification of affected families, greatly facilitated their precise mapping and eventual sequence identification of the culprit mutations. *P53*, which was cloned via a more roundabout protein-based approach, is much more widespread. Paradoxically, the more common and more widespread a tumor suppressor genes is, the more difficult it can be to detect by positional approaches.

New technologies were devised to specifically isolate the DNA sequences that were lost during tumorigenesis. The rationale for this approach was that chromosomal regions that were consistently lost in cancers were likely to contain tumor

suppressor loci. Though this rationale was simple, the technology for comparing cancer cell genomes with their normal cell counterparts was, and is still, laborious. The haploid human genome in total is 3.4×10^9 bp in size; regions of loss can be large and diverse. Methods of high-throughput DNA sequencing and SNP analysis were only beginning to be developed.

An ingenious method to compare cancer and normal cell genomes was developed by Michael Wigler and his colleagues and published in 1993. Termed *representational difference analysis*, this subtractive method allowed the enrichment of lost sequences that were present in one genome but absent in another. In the first step of this complex method, representative regions of both genomes were amplified by PCR. The second step involved iterative cycles of DNA melting, annealing, amplification of rare, hybridized sequences and the degradation of common representations. The final product of this protocol was a small set of short DNAs that were unique to one genomic DNA sample. The analysis of cancer genomes by this approach can amplify regions that were homozygously deleted in cancers. Employed in reverse, to assess genetic gains, representational difference analysis was successfully used to isolate small regions of herpesvirus DNA that are often integrated in the tumor cells of Kaposi's sarcoma, a cancer found in patients with AIDS.

In 1997, Ramon Parsons and his coworkers used a DNA probe derived by representation difference analysis to identify a specific region of loss on chromosome 10. The Parsons group found biallelic loss of their probe sequence in two different breast cancers. The same probe was used to isolate a genomic clone that spanned this homozygous deletion. Sequencing and mapping of the deleted region revealed a previously uncharacterized gene that encoded a 403 amino acid protein. Analysis of the protein sequence revealed several conserved motifs, including a protein tyrosine phosphatase domain and a region to a chicken cytoskeletal protein called tensin. Because of these homologies and the mapping of the gene to chromosome 10, the gene was designated *PTEN*. Independently a collaborative effort by the laboratory of Peter Steck and the company Myriad Genetics found that four brain tumor cell lines that had similarly deleted the same locus, which they designated *MMAC1* for mutated in multiple advanced cancers. In addition to homozygous deletions, the Steck/Myriad group also detected other mutations in prostate, kidney and breast cancers. Finally, a third group, Da-Ming Li and Hong Sun, used similarities shared by protein phosphatase genes, thought to have broad roles in cancer cells, to isolate a gene at 10q23 which they designated *TEP1*. *PTEN, MMAC1* and *TEP1* are all identical; the gene is now most commonly referred to as *PTEN*.

Losses of chromosome 10 sequences had previously been detected by cytogenetic analysis of several types of cancer, including brain, bladder and prostate cancer. LOH analysis was then used to map a common region of loss to chromosome band 10q23. These studies had been highly suggestive of a tumor suppressor gene within a relatively large region that contains many genes. The homozygous deletions located and mapped by the Parsons and Steck/Myriad groups confirmed this prediction. Additionally, it had been determined that the locus for *Cowden disease*, a rare autosomal dominant familial cancer syndrome, was located on chromosome 10. Cowden disease is typified by the presence of benign lesions, called

hamartomas that affect the skin, breast, thyroid, and the oral and intestinal epithelia. Breast and thyroid cancers are also components of Cowden disease. Prior to the cloning of *PTEN*, high-resolution mapping by Charis Eng and her colleagues had demonstrated linkage to the 10q23 region.

The identity of the Cowden disease locus and *PTEN* was soon confirmed. Mutations of *PTEN* were found in over 80% of the Cowden disease families. The mutations found in these families were missense and nonsense point mutations, insertions, deletions, and splice-site mutations, nearly one half of which affected the phosphatase domain at the N-terminus of the encoded protein. The mutated allele was often found to be retained after LOH in tumors, confirming the role of *PTEN* as a tumor suppressor in this syndrome.

Another rare, autosomal dominant disease with clinical features that partially overlaps those of Cowden disease is *Bannayan–Riley–Ruvalcaba* syndrome. Affected individuals with disease develop the benign tumors associated with Cowden disease, but do not typically develop malignancies. Analysis of *PTEN* in Bannayan–Riley–Ruvalcaba syndrome families revealed mutations that segregated with disease. Interestingly, one mutation found in a Bannayan–Riley–Ruvalcaba family was identical to a *PTEN* mutation previously found in a Cowden disease family. This suggests that variable penetrance of *PTEN* mutations can alternatively lead to two clinically distinct syndromes. It is possible that modifier loci might play a significant role in *PTEN* mutation-associated phenotypes.

Overall, 80% of Cowden disease families and 60% of Bannayan–Riley–Ruvalcaba syndrome families have been shown to harbor mutations in *PTEN*. It remains possible that additional mutations in *PTEN* remain undetected or that additional loci play a significant role in these syndromes. The overall incidence of Cowden disease has been estimated to be 1 in 200,000, but the subtle manifestations of the disease and the variable penetrance of *PTEN* mutations suggest that this may be an underestimate.

PTEN is frequently mutated in diverse types of sporadic cancers. The two cancers that most commonly harbor mutated *PTEN* genes are glioblastomas, a type of brain cancer in which up to 45% of tumors are *PTEN*-mutant, and endometrial cancer, in which *PTEN* is mutated in about one half of the samples tested. While germline *PTEN* mutations predispose to breast and thyroid cancers, only about 6% of sporadic forms of these cancers involve *PTEN* mutations. *PTEN* mutations have also been detected in smaller numbers of bladder, ovarian, colon, lymphatic and lung cancers. In some cancer types, *PTEN* mutations are found in a greater proportion of larger, more malignant cancers, suggesting that *PTEN* loss can affect cellular phenotypes related to tissue invasion and motility.

Studies of prostate cancers have revealed that approximately one half exhibit LOH in the 10q23 region, while 10% have defined homozygous deletions at the *PTEN* locus. Similarly, breast cancers also have a high rate of LOH at 10q, while *PTEN* is actually found to be specifically mutated in only about 5% of specimens analyzed. It thus appears to be fairly common that specific *PTEN* mutations are not found in tumors with LOH at 10q23. Why might this be? Similar to *RB, PTEN* appears to be the frequent target of homozygous deletion. This type of mutation can be difficult

to ascertain by routine methods of genetic analysis. Point mutations are typically detected against a background of normal, wild type sequence. In contrast, the absence of signal that arises from attempts to amplify a deleted region can be difficult to quantify and verify. Thus, a lack of complete concordance between LOH and clear evidence of a first 'hit', as predicted by Knudson, is likely the result of technical difficulties inherent in the detection of unequivocal homozygous deletions. Less definitive techniques that measure total *PTEN* expression have shown that expression is commonly reduced in cells with LOH in the 10q23 region, suggesting the retention of a dysfunctional allele. Alternatively these results could suggest that a second, as yet undiscovered, tumor suppressor gene in the 10q23 region may be an alternative target of inactivation in some tumor types.

The protein tyrosine phosphatase activity of the PTEN protein plays a prominent role in the regulation of cell growth and cell death. Interestingly, the proto-oncogene *PIK3CA* also plays an antagonistic regulatory role in this process. How *PTEN* and *PIK3CA* mutations affect the phenotype of the developing cancer cell will be discussed in Chapter 5.

SMAD4 and the Maintenance of Stromal Architecture

Polyps within the gastrointestinal tract occur in 1–2% of children. The majority of these are sporadic lesions of no consequence that are sloughed into the lumen and excreted in the stool. In a small number of individuals, juvenile polyps occur as part of a familial disorder known as *juvenile polyposis syndrome* (JPS). The relationship between JPS-associated polyps and cancer illustrates the relationship between tissue structure, development-associated genes and the risk of malignancy.

The polyps that occur in JPS patients are histologically distinct from the adenomas that are characteristic of FAP. Juvenile polyps are hamartomas, which are focal growths thought to result from faulty developmental processes. Hamartomas within the gastrointestinal tract are composed of a mixture of glandular and stromal elements. Though they resemble neoplasms, hamartomatous polyps grow at the same rate as the normal adjacent tissue and do not invade or otherwise alter the surrounding tissue structure. They are thus more of a structural defect than a growth defect per se. While the adenomas that occur in FAP patients are restricted to the colon and rectum, the hamartomas that occur in individuals affected with JPS occur throughout the upper and lower gastrointestinal tract.

Recognized as an autosomal dominant disorder in 1966, JPS is rare and has an incidence that has been estimated at 1 in 100,000. Though JPS appears to be genetically heterogeneous, linkage to markers on chromosome 18 has been found in approximately one half of known JPS kindreds. Within the interval of linkage on chromosome band 18q21 is a tumor suppressor gene cloned by Scott Kern and his coworkers in 1996.

Allelic loss involving the 18q21 region can be found in about 90% of pancreatic cancers. The Kern group mapped homozygous deletions within this region in a

large panel of sporadic pancreatic carcinomas. These deletions were found to commonly include a locus that they designated *DPC4* (it was the fourth gene that had been reported to be deleted in pancreatic carcinoma). Additional evidence for *DPC4* as a tumor suppressor gene was the finding of inactivating single base substitutions and a small deletion in pancreatic tumors that did not have homozygous deletions. *DPC4* was found to contain significant homology to the *D. melanogaster* Mothers against decapentaplegic (*MAD*) gene and the *C. elegans SMA* gene family, which are all involved in development. *DPC4* became commonly known as SMA- and MAD-related gene 4, or *SMAD4*.

Germline mutations in *SMAD4* have been found in most of the JPS kindreds in which linkage to 18q markers had been established. Among sporadic cancers, *SMAD4* is most commonly inactivated in pancreatic cancers, and in other cancers of the gastrointestinal system. Loss of *SMAD4* function is found in about 15% of sporadic colorectal cancers. It appears that *SMAD4* mutations are uncommon in tumors that occur outside the gastrointestinal tract.

SMAD4 and homologs of *SMAD4* in other species are important regulators of both development and tissue homeostasis. Human *SMAD4* is a member of a *SMAD*-gene family that composes an intracellular communication network. The role of the *SMAD4* encoded protein in this network is to both receive signals communicated from the cell surface and transduce them to the cell nucleus, where gene expression is regulated. The SMAD network is an important mechanism that allows cells to sense changes in their environment, such as those that naturally occur during development and normal growth of tissues, and to orchestrate a measured response to these changes. The role of *SMAD4*-dependent communication in the response of cells to their environment will be described in detail in Chapter 5.

Although hamartomas are benign lesions, the presence of large numbers of hamartomas is a significant risk factor for the development of carcinomas. In the preceding section, we have seen how germline mutations in *PTEN* cause the hamartomatous syndromes Cowden disease and Bannayan–Riley–Ruvalcaba syndrome and a corresponding increase in the risk of many types of cancer. Approximately one half of individuals with Cowden disease have gastrointestinal hamartomas. The extent of clinical overlap between Cowden disease and JPS is significant, and therefore the conclusive diagnosis of JPS largely depends on the exclusion of the other hamartomatous syndromes. Correctly categorizing and diagnosing the patient with gastrointestinal hamartomas is challenging, but important. Individuals with Cowden disease must be monitored carefully for the development breast and thyroid cancers, while JPS does not carry these risks. In the near future, the differing genetic basis for these syndromes will be a useful tool for specific diagnosis and risk analysis.

Until relatively recently it was unclear whether gastrointestinal hamartomas actually develop into carcinomas, or whether gastrointestinal cancers arise from distinct precursor lesions in the same patients. Analysis of large numbers of sporadic and inherited juvenile polyps have revealed regions of adenomatous epithelium in a small proportion of these lesions. It thus appears that each hamartoma associated with Cowden disease or JPS has the potential, albeit low, to progress to a carcinoma.

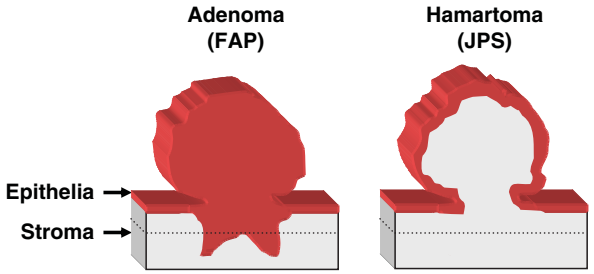

Fig. 3.13 Two types of colorectal polyps. In patients with FAP, germline mutations in *APC* lead to the development of hundreds of adenomas. Adenomatous polyps are composed primarily of epithelia (red). Mutant epithelial cells carry a significant risk of further clonal evolution. The hamartomatous polyps characteristic of JPS are caused by germline mutations in *SMAD4*. In contrast to adenomas, hamartomas are composed primarily of stroma (gray). Stromal cells do not themselves evolve into cancers, but their proliferation alters the landscape of the colon epithelium. The resulting changes in the microenvironment provide selective pressure for the outgrowth of epithelial neoplasia

The adenomatous polyps associated with FAP are largely composed of epithelial cells. Analysis of adenoma cells has revealed clonal genetic defects that are associated with tumor progression (Chapter 1). In contrast with the adenomatous polyps of FAP, the hamartomas associated with JPS are composed largely of stromal cells (see Fig. 3.13). Genetic losses have been detected in these stromal growths, suggesting that they are expanded clones. However, JPS does not predispose affected individuals to stromal cell cancer. The cancer associated with JPS is colorectal carcinoma, which, like all carcinomas, arises from epithelial tissue. The conclusion that can be drawn from these findings is that genetically mediated changes in the stroma can create an environment that promotes the outgrowth of epithelial cell clones, which progress to cancers.

The induction of tumors by the alteration of the stromal environment represents a distinct mechanism of tumorigenesis. Histological examination of hamartomas shows that epithelial cells become entrapped within abnormal stroma. These entrapped epithelial elements form dilated cysts and develop areas of local inflammation. As described in Chapter 1, inflammation provides a microenvironment in which the clonal growth of cancer cell precursors can be selected. Thus, early genetic changes that occur in stromal cells can predispose neighboring epithelial cells to grow into tumors.

Two Distinct Genes Underlie Neurofibromatosis

Neurofibromatosis is a genetic disease that is characterized by numerous benign, lesions. As in the case of *SMAD4*, the mutations that cause neurofibromatosis lead to defects in tissue architecture and simultaneously cause a predisposition to cancer.

Fig. 3.14 Neurofibromatosis type 1. Severe disease is apparent on the torso of a 45-year-old woman. Café-au-lait macules (straight arrows) and neurofibromas (curved arrows) are indicated. (From Cohen, P. R. *New Engl. J. Med.* **329**, 1549 (1993).) (Copyright 1993 Massachusetts Medical Society. All rights reserved.)

The genetic alterations that cause neurofibromatosis are particularly devastating to affected individuals because the characteristic lesions are externally evident (see Fig. 3.14).

The most common form of neurofibromatosis is Neurofibromatosis 1 (NF1), also known as Van Recklinghausen neurofibromatosis. Affected individuals exhibit pigmented lesions known as café-au-lait spots, freckling and hamartomas in the iris of the eye known as Lisch nodules. NF1 is strongly associated with cognitive dysfunction, including mental retardation and learning disabilities. In addition to the diagnostic, disabling, features of the disease, patients affected by NF1 are prone to unusual malignancies. A common feature of such cancers is that they occur in

tissues that developmentally arise from the neural lineage. NF1 patients develop tumors in the sheath of peripheral nerves (neurofibrosarcomas); such tumors are highly aggressive and metastatic. NF1 is also strongly associated with tumors in the optic nerve (optic gliomas) which rarely become symptomatic. Cumulatively, between 2% and 5% of patients affected by NF1 develop cancer, a rate that is significantly higher than that in the general population.

The NF1 gene was cloned by the combined use of physical mapping and linkage mapping. A large-scale mapping effort used data derived from 142 families with over 700 affected individuals to localize the gene to17q11.2. This effort was accelerated by the analysis of two patients, in whom balanced translocations with defined break points narrowed the search considerably. Candidate genes from the narrowed region were evaluated by DNA sequencing. A large gene, spanning 300 kb and containing a 9 kb open reading frame was identified independently by groups led by Francis Collins and Ray White, and reported in 1990.

Mutations in the NF1 gene, designated *NF1*, include large deletions, small rearrangements, and most frequently, point mutations. The latter type of mutation is distributed throughout the *NF1* coding sequences. Unlike some common tumor suppressor genes, *NF1* does not contain mutation hotspots. The protein encoded by *NF1*, called neurofibromin, bears significant homology to a family of signaling proteins that regulate cell size, shape and proliferation. The relationship of *NF1* to cell signaling pathways involved in cancer is described in Chapter 5.

With a prevalence in the population that is estimated at 1 in 3,000, NF1 is one of the most common autosomal dominant disorders in humans. The majority of NF1 cases are inherited, but it appears that a significant proportion of cases arise from newly arising germline mutations. Single mutated *NF1* alleles are dominant, and are sufficient to cause the clinical manifestations of NF1. During the development of NF1-associated cancers, the remaining wild type tumor suppressor allele appears to be lost via LOH, as in other syndromes of cancer predisposition. In addition to its role in neurofibromatosis, the *NF1* gene has been found to be mutated infrequently in sporadic tumors arising in tissue of neuroectodermal lineage, including melanomas, neoroblastomas, pheochromocytomas, and neurofibrosarcomas.

A second form of neurofibromatosis is called NF2 or central neurofibromatosis. NF2 is clinically distinct; affected individuals exhibit retinal hamartomas, but do not have the other lesions associated with NF1. NF2 patient frequently develop bilateral tumors of the eighth cranial nerve known as vestibular schwannomas, as well as benign tumors that affect the central and peripheral nervous system. NF2 is considerably less common than NF1 and accounts for about one tenth of neurofibromatosis cases. Many NF2 cases occur in the absence of parental involvement. As is the case with *NF1*, new germline *NF2* mutations appear to arise frequently.

The gene that causes NF2 was cloned independently in 1993 by groups led by James Gusella and Gilles Thomas. *NF2* mutations are found in the germline of NF2 patients and also in sporadic schwannomas, indicating that *NF2* functions as a true tumor suppressor. *NF2*-associated schwannomas rarely develop into malignant lesions, and the overall rate of cancer in NF2 patients is not significantly increased over that in the general population.

Multiple Endocrine Neoplasia Type 1

There are several disorders of cancer predisposition that are characterized by the occurrence in individual patients of multiple cancers that arise in endocrine tissues. Such diseases are termed *multiple endocrine neoplasias*. Within this broad category are several distinct diseases that arise as a result of known genetic alterations. As is the case with NF1 and NF2, the multiple endocrine neoplasias represent cancer predisposition syndromes that are most often inherited but which can apparently also arise sporadically via the appearance of new germline mutations.

There are two forms of multiple endocrine neoplasia that have been well described at the genetic level. The disease caused by inheritance of the *RET* oncogene (see Chapter 2) is termed multiple endocrine neoplasia type 2. In contrast, multiple endocrine neoplasia type 1 (MEN1) is caused by the mutation of a tumor suppressor gene, *MEN1*.

MEN1 is most commonly characterized by tumors in the parathyroid glands and the anterior pituitary gland, neuroendocrine tumors in the pancreas, and carcinoid tumors in the gastrointestinal tract. The latter two types of tumors arise in tissues that are related developmentally to tissues of ectodermal origin. MEN1 often occurs simultaneously with Zollinger–Ellison syndrome, a disorder caused by gastrin-secreting tumors of the pancreas and duodenum.

The *MEN1* gene was localized to chromosome 11 and cloned in 1997 by Stephen Marx and colleagues. Sequence analysis revealed heterozygous inactivating mutations of *MEN1* in individuals with MEN1; the wild type copy of *MEN1* is subsequently lost via LOH during tumorigenesis. LOH of the *MEN1* locus at 11q13 is also frequently seen in sporadic endocrine tumors. The extent to which *MEN1* is mutated in sporadic tumors has not been thoroughly documented. The cellular function of the *MEN1*-encoded protein, menin, is unknown.

The overall incidence of MEN1 is estimated to be approximately 1 in 70,000. This figure is based on clinical, rather than genetic criteria, and may therefore be subject to ascertainment biases. The majority of multiple endocrine neoplasia cases are familial, but a significant number of cases appear to occur spontaneously, in the absence of a family history of endocrine tumors. Unlike neurofibromatosis, multiple endocrine neoplasia cannot be diagnosed in the absence of tumors. Thus, it is possible that some of the cases that appear sporadic may in fact reflect unrecognized familial disease.

Most Tumor Suppressor Genes are Tissue-Specific

Tumor suppressor gene inactivation is generally both tissue- and cancer-specific. As we have seen, mutations in *APC* strongly predispose to carcinomas in the colorectal mucosa. FAP patients with mutant *APC* are not predisposed to lung cancer or breast cancer, even though these malignancies similarly arise in epithelial cell populations. Loss of APC function provides a selective advantage for the outgrowth

of colorectal adenomas, but does not appear to lead to precancerous lesions in other tissues. The underlying reason behind this tissue specificity is not apparent. Two factors that are likely to be related to specificity are the unique cellular architecture of colorectal crypts and the surrounding stroma, and the mechanism by which crypts are continually renewed. Presumably these distinctive characteristics of the large bowel somehow cause a reliance on APC protein activity for the maintenance of homoeostasis that is not present in other epithelial tissues.

In some cases, the tissue compartment in which a given tumor suppressor gene is required to repress neoplastic growth can be precisely delineated. For example, tumors that arise as a result of the biallelic inactivation of *NF2* appear to be largely restricted to the nerve sheath that surrounds the eighth cranial nerve.

The relationship between distinct tumor suppressor genes and specific tumors also extends to those genes that function in connected biological pathways. As will be described in Chapter 5, RB and p16 proteins function in a common molecular pathway that regulates the progression of the cell cycle. This growth controlling pathway can be disrupted by mutation of either *RB* or *CDKN2A*. Nonetheless, the types of tumors in which each of these two genes are found to be mutated do not overlap. Why does the same pathway tend to be disrupted by *CDKN2A* mutations in melanomas, but by *RB* mutation in retinoblastomas? This interesting question remains unanswered at present. It is likely that these genetic losses, though they might affect different points of the same pathway, are not completely functionally equivalent.

Another revealing observation arises from the comparison of tumor suppressor gene mutations in inherited and sporadic forms of cancer. Tumor suppressor genes that are mutated in familial cancer syndromes, such as *APC* in FAP, are often mutated in sporadic cancers involving the same tissues (see Table 3.1). However, this is not always the case. *P53* is mutated in a large proportion of colorectal cancers, but the germline *P53* mutations that cause Li Fraumeni syndrome do not predispose to colorectal cancer. Conversely, germline *BRCA1* and *BRCA2* mutations cause hereditary breast cancer but these genes are not mutated in a significant proportion of sporadic breast cancers. While *SMAD4* was cloned on the basis of its loss in sporadic pancreatic cancers, germline mutations of *SMAD4* have not been found in known familial clusters of pancreatic cancer. Rather, as we have seen, mutated *SMAD4* is found in a subset of JPS kindreds. One general conclusion that can be drawn from these observations is that the same gene may contribute to different forms of cancer in distinct ways. In some cases, mutation of a given tumor suppressor genes may be not absolutely required for the development of a cancer, but later inactivation may nonetheless contribute to later stages of growth and invasion.

Modeling Cancer Syndromes in Mice

Our understanding of tumor suppressor genes and their functions has been confirmed and expanded by studies of genetically engineered mice. The effects of inheriting cancer-associated mutations can be recapitulated in the mouse by the

manipulation of the mouse germline. In general, targeted disruption of tumor suppressor genes in mice causes a significantly increased rate of cancers. Such cancer-prone mice provide valuable model systems for studying gene function and for the preclinical testing of new cancer therapies and preventive agents.

Genes can be disrupted in mouse embryonic stem cells by a process known as *gene targeting*. Briefly, a DNA construct containing an altered gene is transferred to cultured embryonic stem cells. In a small proportion of cells, this construct will integrate into the homologous chromosomal locus and disrupt the gene under study. These modified stem cells are injected into mouse embryos. A small proportion of the chimeric embryos will incorporate the modified stem cells into germ cells during subsequent development, allowing the modified gene to enter the germline of the new strain. Animals with heterozygous disruptions are interbred to achieve homozygosity at the desired locus. A strain of mice with a heterozygous or homozygous loss of a gene by this gene targeting approach is known as a *knockout*. The specifics of this approach are described elegantly in a number of useful texts.

Knockout mice are extremely powerful tools because they allow the effects of loss-of-function mutations to be directly assessed in an intact animal model. In many cases, tumor suppressor gene knockouts result in dramatic phenotypes. Knockout mice that are homozygous for *P53* null alleles develop tumors by the age of 9 months and typically succumb to cancer several months well before 1 year of age. Spontaneous cancers are rare in laboratory strains of mice that have wild type *P53* alleles, and these mice typically have a lifespan of 2–3 years.

Heterozygous *P53* knockout mice genetically model Li Fraumeni syndrome. These mice are also cancer-prone but show a longer latent period prior to cancer development and longer survival as compared with homozygous *P53* knockout mice. The cancers that develop in heterozygous *P53* knockout mice are primarily sarcomas and lymphomas. Only a small proportion of *P53* heterozygous knockouts develop carcinomas, which are the types of cancer that develop most frequently in humans with Li Fraumeni syndrome. The spectrum of tumors that develop in humans and mice with a single functional *P53* allele thus overlaps, but is not identical.

Heterozygous carriers of inactivating *APC* mutations develop intestinal polyposis. As in humans, these polyps exhibit LOH of the *APC* locus, with retention of the mutant *APC* allele. Many of these polyps become cancerous. Interestingly, in *APC* heterozygous knockout mice, the majority of polyps occur in the small intestine rather than the colon.

Compound knockouts, in which two or more genes are simultaneously altered, can be particularly informative. Inactivation of one *SMAD4* allele in mice does not lead to an increased rate of tumors. However, homozygous targeting of both *SMAD4* and *APC* leads to more rapid progression of tumorigenesis in mice than is observed with *APC* targeting alone. This finding supports the human data suggesting that *SMAD4* loss of function in stromal cells increases the rate at which epithelial cancers can arise.

In the case of *CDKN2A*, mouse models have provided answers but also posed additional questions. The design of the initial *CDKN2A* knockout mouse strain

effectively eliminated the expression of both *CDKN2A*-associated transcripts; these mice exhibited a clear cancer-prone phenotype. The discovery of the p14(ARF) transcript in human cells led to subsequent attempts to specifically target the p14(ARF) mouse homolog, a transcript that encodes a somewhat larger protein known as p19(ARF). It was found that knockouts that eliminated p16 but retained p19(ARF) expression were cancer-prone. However, the mouse knockout that eliminated p19(ARF) but retained p16 expression was similarly prone to cancer. The conclusion of these experiments is that, in mice, the genetic elements that encode both p16 and p19(ARF) are independently of critical importance in tumor suppression.

What is the meaning of this result? In humans, the region that uniquely encodes p14(ARF) has not been found to be mutated, either in sporadic tumors or in the germlines of cancer-prone individuals. To date, all of the validated mutations that affect p14(ARF) also involve p16. It is certainly possible that human mutations affecting only p14(ARF) remain to be discovered. Alternatively, it is possible that the roles of the two *CDKN2A* transcripts are different in humans and mice. The p14(ARF) open reading frame does not appear to be evolutionarily conserved. At the sequence level, human p14(ARF) and murine p19(ARF) share only 50% identity. However, there is evidence that some functions of p19(ARF) are conserved in p14(ARF). Based on the mouse data alone, should p14(ARF) be considered a human tumor suppressor gene?

In cases where the data from human cancers and mouse models appear to conflict, it is imperative to prioritize sources of information. From a biomedical standpoint, mouse models are important only in that they provide insight into human cancer. The p19(ARF) knockout mouse model clearly suggests that human p14(ARF) might be a tumor suppressor gene. If human mutational data consistently fail to conform to this prediction, then p19(ARF) will remain a curious mouse transcript that is primarily of interest to those who study developmental and evolutionary biology.

The results of these exemplary studies illustrate both the unparalleled strengths and limitations of mouse cancer models. Knockout mice have confirmed the hypothesis that inactivation of tumor suppressor genes are critical and rate limiting events during tumorigenesis. While mutated tumor suppressor genes clearly cause mice to be prone to cancer, the differences between human cancer syndromes and the phenotypes of knockout mice are significant. Perhaps this is not surprising at all, given the high degree of divergence between the two species. In humans, cancers and particularly carcinomas, are strongly associated with aging. Among many other differences, mice and humans have dramatically different lifespans. The relatively short lifespans of the mouse may partially explain the relative paucity of carcinomas in *P53* knockout mice. Many genetic factors also undoubtedly contribute to the divergent phenotypes observed. Indeed, *P53* knockout mice in different genetic backgrounds can exhibit clearly distinct cancer phenotypes. Thus, despite some limitations, mouse models clearly illustrate both the simple and the complex principles of cancer genetics.

Tumor Suppressor Gene Inactivation During Colorectal Tumorigenesis

How do losses of tumor suppressor gene functions contribute to tumor development? The timing of common genetic losses during tumorigenesis provides considerable insight. The most comprehensive genetic model for multistep tumorigenesis is based upon extensive data collected from colorectal cancers and their precursor lesions. In the colorectal mucosae, characteristic genetic losses demarcate the stages of tumor growth (see Fig. 3.15).

Inactivation of *APC* is a very frequent event in both inherited and sporadic forms of colorectal cancer. *APC* mutants are thus highly prevalent in this type of cancer. Inherited mutants are highly penetrant. Even in the absence of other data, these observations strongly suggest that *APC* inactivation is a rate-limiting step in colorectal tumorigenesis. Further insight can be gained from examination of lesions at different stages. Mutations of *APC* and losses of chromosome 5q are found in the entire spectrum of colorectal neoplasia, from small adenomas to metastatic cancers. *APC* mutations are found in the majority of each of these lesions. As will be described in Chapter 5, the small proportion of colorectal tumors that have wild

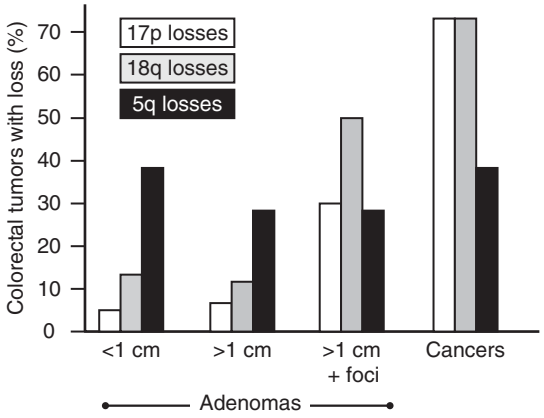

Fig. 3.15 Chromosomal losses during colorectal tumorigenesis. There is a high frequency of LOH involving chromosomes 17p, 18q and 5q in colorectal adenomas and invasive cancers. Allelic losses involving 17p (that contain the *P53* locus) and 18q (that contain the *SMAD4* locus) tend to occur predominately in larger adenomas that contain focal regions of carcinomatous transformation, and in cancers. In contrast, allelic losses of chromosome 5q sequences (that contain the *APC* locus) occur at similar frequency in small adenomas, larger adenomas and cancers. These data suggest that 5q loss is an early event, while 17p and 18q losses occur later in tumorigenesis. Note that an evaluation of large allelic losses can underestimate or overestimate of the extent of tumor suppressor gene inactivation. Smaller deletions and other mutations are not detected by this type of analysis, while large regions of loss can involve multiple tumor suppressor genes. (Data from *The Genetic Basis of Human Cancer*, Kinzler & Vogelstein, eds., McGraw-Hill (2002).)

type *APC* often contain mutations in *CTNNB1*, a gene that functions in concert with *APC*. Even the earliest lesions analyzed, aberrant crypt foci, have been found to harbor *APC* mutations. The unifying model derived from these observations is that functional inactivation of *APC* triggers the first waves of clonal expansion of cancer precursors.

The pattern of *P53* inactivation in colorectal cancers is different from that of *APC*, suggesting a distinct role for these two events. While *P53* mutations and losses involving the *P53* locus on chromosome 17p are frequently, but not always, found in advanced colorectal cancers, they are found much less frequently in precursor lesions (see Fig. 3.15). *P53* inactivation is therefore a relatively late event in colorectal tumorigenesis. Carriers of *P53* mutations (individuals affected by Li Fraumeni syndrome) do not appear to be at increased risk of colorectal cancer. Collectively, these observations suggest that *P53* inactivation does not initiate the process of colorectal tumorigenesis, but rather plays a role in the transition from larger adenomas to invasive cancers.

Frequently occurring after the early inactivation of *APC* is allelic loss on 18q involving the *SMAD4* locus. Losses of 18q are frequently seen in large (>1 cm), late stage adenomas and in invasive cancers. These alterations are rarely observed in less advanced lesions. Thus, the inactivation of tumor suppressor loci on 18q typically occurs during intermediate stages of tumor progression. Many cancers with 18q losses also exhibit mutation of *SMAD4*, indicating that *SMAD4* is likely the target of inactivation in these cancers. However, the fact that the overall frequency of *SMAD4* inactivation (~15%) is significantly lesser than the frequency of 18q loss (>50%) suggests that inactivation of additional tumor suppressor loci in the 18q region, apart from *SMAD4*, probably also plays a role during the intermediate stage of some colorectal tumors.

Colorectal cancer provides a useful model for understanding how cancers arise and progress in step with accumulating genetic alterations (see Fig. 3.16). While the genetic principles learned from analysis of colorectal cancers appear to be generally applicable to all cancers, the specific genes involved and the roles they play can vary. As we have seen, many genetic alterations are tumor specific. *APC* mutations are ubiquitous in colorectal cancers, but generally not observed outside of the gastrointestinal tract. Presumably, other cancer types have a gene, or perhaps several genes that can play a similar role as *APC* in the initiation of tumors.

The genes that play defined roles in colorectal cancer might play somewhat different roles in cancers arising from other tissues. *P53*, for example, is inactivated in a very broad spectrum of human cancers. The available evidence suggests that *P53* inactivation is likely to occur at different stages in different cancers. Patterns of *P53* loss in sporadic and inherited breast carcinomas, for example, suggest that *P53* loss may be a rate-limiting step in the development of this type of cancer. Among women affected by Li Fraumeni syndrome, who are heterozygous for inactivating *P53* mutations, breast carcinomas are the most common cancers. *P53* mutations and chromosome 17p losses can be found in many sporadic breast cancers and also in noninvasive precursor lesions.

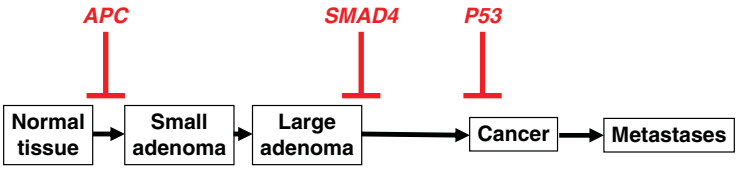

Fig. 3.16 Tumor suppressor genes define rate-limiting steps in colorectal cancer evolution. The combination of LOH and mutational data support defined roles of tumor suppressor genes. *APC* controls the rate-limiting step of initial adenoma formation. The selection for *SMAD4* and *P53* losses occur later in the process of tumorigenesis, as larger adenomas evolve into invasive cancers

Inherited Tumor Suppressor Gene Mutations: Gatekeepers and Landscapers

In previous sections of this chapter, we have seen how the inheritance of germline tumor suppressor gene mutations leads to an increased risk of cancer. As seen in Table 3.1, these risks – measured as the allele penetrance – vary considerably. The variable penetrance of different tumor suppressor gene mutations reflects the distinct ways in which these alleles contribute to tumor development. This principle can be exemplified by comparing the rates of colorectal cancer associated with germline mutations in *APC* and *SMAD4*.

Inactivated *APC* alleles are highly penetrant while inactivated *SMAD4* alleles are less so. Why do different types of mutations confer different risks for the same disease? The answer to this question lies in the effect of a genetic loss on the phenotype of a tumor cell, and whether this effect is direct. The most potent tumor suppressor genes have direct effects on cell growth. Mutations of growth-controlling genes are typically highly penetrant and thus confer the greatest risk of cancer. Inactivation of *APC* directly causes the outgrowth of pre-cancerous polyps. In FAP, every cell has only one functional *APC* allele. Loss of this single allele is sufficient to give rise to a polyp. The large number of polyps that occur in FAP patients is a virtual guarantee that some will eventually develop into cancers. One can readily infer that wild type *APC* must play a critical role in regulating cell growth and preventing neoplasia.

Classical growth-controlling tumor suppressor genes such as *APC* have been categorized as '*gatekeepers*'. Gatekeepers directly suppress cell outgrowth. Cells that lose gatekeeper activity form neoplasia, each of which has the potential to become a cancer. When wild type gatekeeper genes are experimentally reintroduced into established cancer cells, they typically lead to suppression of growth.

The inherited mutations of *SMAD4* affect epithelial cell populations in a less direct manner. Germline *SMAD4* mutations appear to primarily alter the growth of stromal cells that are not cancer precursors. *SMAD4* inactivation thus alters the tissue structure of the colorectum. This abnormal microenviroment provides a fertile landscape for the outgrowth of epithelial neoplasia. Mutations in *SMAD4* typify what has been termed a '*landscaper*' defect.

Defects in the tissue landscape can also be caused by chronic inflammation. The disease ulcerative colitis is a premalignant condition characterized by inflammation of the wall of the bowel and an increased risk of colorectal cancer (see Chapter 1).

RB clearly functions as a gatekeeper in the cells of the developing retina. In contrast, the timing of *P53* inactivation in later stages of colorectal cancer suggests that *P53* does not function as a gatekeeper in the colorectal epithelium. However, a significant body of evidence suggests that *P53* is a gatekeeper in other cancer types, most notably breast cancers.

Maintaining the Genome: Caretakers

A third category of tumor suppressor genes affect cancer precursor cells directly, but not by controlling their growth. Rather, the proteins encoded by these genes function to maintain a stable genome by directly participating in various processes of DNA repair. When DNA repair protein-encoding genes are inactivated, the overall rate of mutation increases. All subsequent generation of cells then have an increased tendency to inactivate additional tumor suppressor genes and to activate oncogenes. The process of tumorigenesis is thus accelerated. The genes that function to maintain genetic stability are known as '*caretakers*'. Examples of caretaker genes are *BRCA1* and *BRCA2*, breast cancer susceptibility genes that are required for DNA repair (see Chapter 5).

Caretaker defects define a unique category of tumor suppressor genes. As we have seen in the case of familial breast cancers, the penetrance of mutated caretakers varies considerably. Because they function in the repair of DNA lesions, caretakers are an intrinsic component of the cellular response to mutagens in the environment. Accordingly, environmental factors have a significant role in determining the penetrance of caretaker gene mutations.

A caretaker defect is the defining characteristic of an inherited colorectal cancer syndrome called hereditary nonpolyposis colorectal cancer (HNPCC). Defects in one of a family of genes involved in a specific DNA repair process cause an overall increase in the rate of somatic mutations. HNPCC illuminates the central role of genetic instability in cancer and will be described in detail in Chapter 4.

Further Reading

Collins, F. S. Positional cloning moves from perditional to traditional. *Nat Genet.* **9**, 347–350 (1995).
de la Chapelle, A. Genetic predisposition to colorectal cancer. *Nat. Rev. Cancer* **4**, 769–780 (2004).
Dyer, M. A. & Bremner, R. The search for the retinoblastoma cell of origin. *Nat. Rev. Cancer* **5**, 91–101 (2005).
Frese, K. K. & Tuveson, D. A. Maximizing mouse cancer models. *Nat. Rev. Cancer* (2007).
Kinzler, K. W. & Vogelstein, B. Lessons from hereditary colorectal cancer. *Cell* **87**, 159–170 (1996).

Kinzler, K. W. & Vogelstein, B. Landscaping the cancer terrain. *Science* **280**, 1036–1037 (1998).
Marx, S. J. Molecular genetics of multiple endocrine neoplasia types 1 and 2. *Nat. Rev. Cancer* **5**, 367–375 (2005).
Narod, S. A. & Foulkes, W. D. BRCA1 and BRCA2: 1994 and beyond. *Nat. Rev. Cancer* **4**, 665–676 (2004).
Sharpless, N. E. & DePinho, R. A. The INK4A/ARF locus and its two gene products. *Curr. Opin. Genet. Dev*. **9**, 22–30 (1999).
Soussi, T., Ishioka, C., Claustres, M. & Beroud, C. Locus-specific mutation databases: Pitfalls and good practice based on the p53 experience. *Nat. Rev. Cancer* **6**, 83–90 (2006).

Chapter 4
Genetic Instability and Cancer

What is Genetic Instability?

When a cell divides, its genome is first duplicated and then distributed to each daughter cell. Every aspect of this fundamental biological process is tightly controlled, ensuring that the information encoded in the genomic DNA does not significantly change as it passes from generation to generation. A full complement of chromosomes is inherited in structurally intact form. The process of DNA replication is similarly characterized by an extraordinarily high degree of fidelity. During the proliferation of normal cells, heritable genetic changes occur only rarely. The information content of the genome in the cells that compose normal tissues is highly stable over the lifetime of the individual.

Cancer cells exhibit defects in the mechanisms by which the genome is replicated and repaired and by which chromosomes are segregated. Not all of these defects are present in every cancer cell, but it appears that every cancer cell has at least one of these types of defects. The result is that the rate at which genetic alterations occur is consistently higher in cancer cells than in normal proliferating cells. Genetically, the cells of a growing tumor are significantly less stable than those in neighboring normal tissues.

Why is the genetic instability exhibited by tumor cells important? As we have seen in the preceding chapters, tumors are caused by sequential genetic alterations. These genetic alterations do not arise all at once but coincide with each wave of clonal expansion that defines a stage of tumor development. An increase in genetic instability means that the cells of a developing tumor will acquire genetic alterations at a greater rate than would otherwise be expected. The genetic instability that occurs during the process of tumorigenesis immediately serves to accelerate the occurrence of all subsequent genetic alterations. Put another way, genetic instability increases the pace of clonal evolution.

Genetic instability is a heritable cellular phenotype. The genetic instability observed in a cancer cell is the result of an ongoing defect in genome maintenance or chromosome transmission. This is a concept that is of singular importance in cancer genetics. A frequent point of misunderstanding is the relationship between genetic instability, which is a defect in a process, and genetic alterations, which are stochastic events. A random mutation does not necessarily indicate, nor cause, genetic instability.

F. Bunz, *Principles of Cancer Genetics*.
© Springer 2008

As we have seen in the preceding chapters, mutations that inactivate tumor suppressor genes and activate oncogenes can be found in all cancers, demonstrating that they are not merely incidental occurrences but central defining features of cancer cells. Similarly, all cancers exhibit a form of genetic instability. The precise type of genetic instability and the mechanism by which these instabilities cause increased rates of genetic alterations may vary in different cancers. Nevertheless, some form of instability appears to be associated with every type of cancer. Genetic instability is thus a defining characteristic of cancer cells.

The Majority of Cancer Cells are Aneuploid

One of the most readily observable traits of cancer cells is an excess number of chromosomes. While normal somatic cells invariably contain 23 pairs of chromosomes, the cells that compose tumors often deviate significantly from this diploid complement. A cell that has a number of chromosomes that is not a multiple of the haploid number is defined as *aneuploid*. Aneuploid cancer cells typically contain between 60 and 90 chromosomes, and this number varies from cell to cell within a single tumor.

In addition to these numerical abnormalities, the chromosomes in aneuploid cells commonly have structural aberrations that are rarely observed in normal cells. The structural abnormalities associated with aneuploidy include translocations, deletions, inversions and duplications.

When observed during mitosis, aneuploid cells exhibit mechanical defects in chromosome segregation. The features of aneuploidy in cancer cells were first described by David Hansemann a decade after the discovery of chromosomes in the late 1870s. Upon microscopic examination of carcinomas, Hansemann observed several recurring chromosomal abnormalities. Prominent among these were asymmetrical mitotic figures that appeared to result in ‘imbalances’ in the chromosome complement of daughter cells (see Fig. 4.1). While abnormal mitoses and chromosome complements had been observed in cancer tissues before, the prevailing thinking had held that these features were the result of fusions between neighboring tumor cells. Contrary to this idea, Hansemann proposed that the observed defects in what is now called *chromosome segregation* were an intrinsic defect in cancer cells and a causative factor in tumorigenesis. The hypothesis put forth by Hansemann was extended and popularized several years later by Theodor Boveri, who emphasized the fact that mitotic spindles in cancer cells were often multipolar (see Fig. 4.1), suggesting that an underlying mechanical defect was an essential feature of the cancer cell.

Modern cytogenetic techniques vividly reveal both the numerical and structural abnormalities that define aneuploidy (see Fig. 4.2). While they are highly illustrative, karyotypes such as those shown significantly underestimate the true extent of sequence gains and losses. The reason for this disparity is that cytogenetic techniques can mark chromosomes, but cannot distinguish submicroscopic changes,

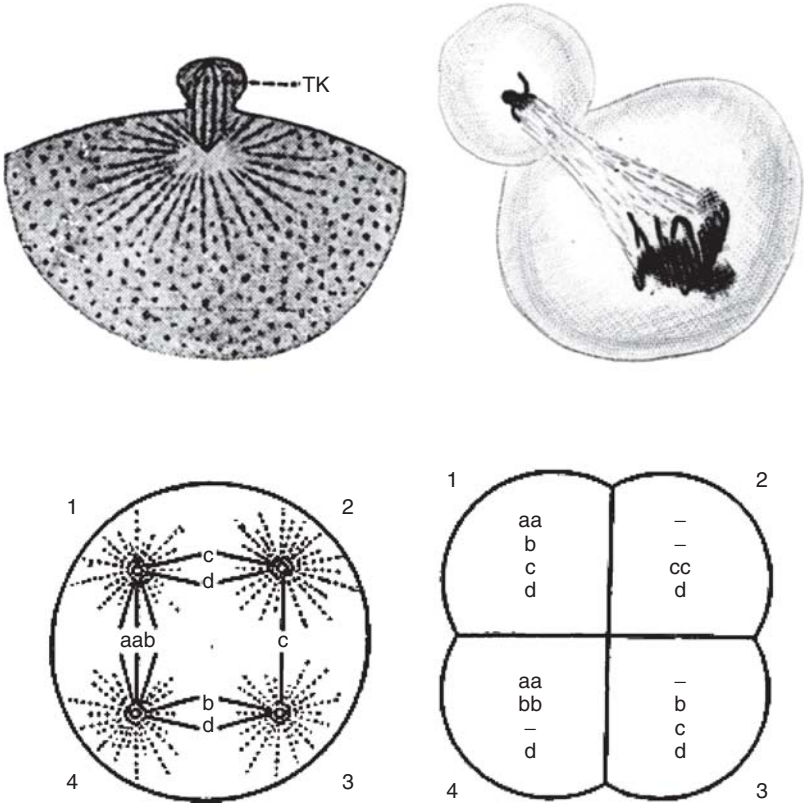

Fig. 4.1 Early observations of aberrant mitoses in cancer cells. An asymmetrical mitotic figure (top panel; from Hansemann. *Virschows Arch. Pathol. Anat.* **119,** 299–326 (1890)) and a tetrapolar mitotic figure (bottom panel; from Boveri *Zur Frage der Entstehung maligner Tumoren*, Gustav Fischer Verlag, Jena (1914))

such base changes. Molecular techniques such as SNP analysis can make this distinction. As an example, consider a cell that has lost a maternal chromosome 17, but then reduplicated the corresponding paternal chromosome 17. This cell would have a normal karyotype, but would have lost every unique allele carried on the maternal chromosome 17. The use of molecular techniques has revealed that in many common cancers, 25% of alleles are lost, while losses of greater than half of all alleles are not unusual.

What is the meaning of these striking cellular perturbations? Is aneuploidy causally involved in cancer, or merely an effect of cancerous growth? This point has been a matter of vigorous debate in the century that has elapsed since the observations of Hansemann and Boveri. The prevalence of aneuploidy in cancer would suggest that it contributes to the process of tumorigenesis, but it has also been proposed that aneuploidy is merely a byproduct of dysregulated cell growth

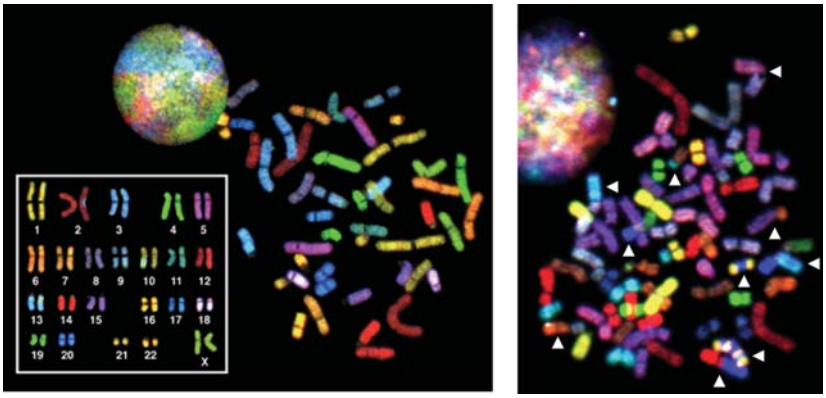

Fig. 4.2 Spectral karyotyping. With the use of chemical inhibitors of mitotic spindle formation, cultured cells can be blocked in metaphase, facilitating the examination of individual, condensed chromosomes. After fixation, these cells are incubated with chromosome-specific DNA probes that are conjugated with fluorophores. The hybridization of these probes with fixed chromosomes effectively results in the painting of each chromosome with an identifiable color. The cell analyzed in this example has a diploid chromosome complement, with no gross structural abnormalities. (left panel; image courtesy of NHGRI). A spectral karyotype of a cancer cell reveals both numerical and structural abnormalities (right panel). Note the numerous chromosomal rearrangements (indicated by arrowheads). (Courtesy of Constance Griffin, MD, Johns Hopkins University.)

or structural changes that arise during tumorigenesis. In the sections that follow, we will explore the relationship between aneuploidy and the cancer gene theory.

Aneuploid Cancer Cells Exhibit Chromosome Instability

The descriptions of aneuploidy provided by Hansemann and Boveri suggested that aneuploidy might be a manifestation of an underlying defect in mitosis. An alternative interpretation is that aneuploidy arises by some other means, and that mitosis is simply more likely to fail in the presence of too many chromosomes. A powerful approach to testing these two possibilities was devised by Christoph Lengauer, while working with Bert Vogelstein and Kenneth Kinzler in the late 1990s. Using the technique of fluorescence in situ hybridization, Lengauer measured the rates at which chromosomes are lost and gained in colorectal cancer cells during long-term culture (see Fig. 4.3). Diploid cancer cells were observed to maintain a stable chromosome complement when propagated for many generations. In contrast, cancer cells that were aneuploid tended to gain and lose individual chromosomes at the relatively high rate of 0.01 per chromosome per cell division. When clonal populations of these aneuploid cells were propagated for multiple generations, the cells within each clone were found to rapidly diverge from one another with respect to

their chromosome complement. The increased rate of chromosome gains and losses in aneuploid cells was termed *chromosomal instability*, or CIN.

Additional insight was gained by cell fusion experiments. Hybrid cells resulting from the fusion of two diploid cells maintained a constant chromosome number, despite the fact that these cells contained an aberrant chromosome complement. Thus, one feature of aneuploidy could be experimentally separated from the underlying process that causes CIN. Fusions between diploid and aneuploid cells resulted in cells that were CIN. Several conclusions can be drawn from these experiments: (1) aneuploidy is a reflection of an ongoing cellular process, (2) aneuploidy does not *cause* instability, but rather may *result* from instability, and (3) CIN is a dominant phenotype.

The quantification of CIN provides a useful framework for understanding the nature of aneuploidy and its potential role in cancer. Aneuploidy is a state that reflects an ongoing, dynamic process which can be measured as CIN. A significant body of evidence, including the original observations of Hansemann and Boveri, suggest that the mitotic defects observed in aneuploid cells contribute to CIN.

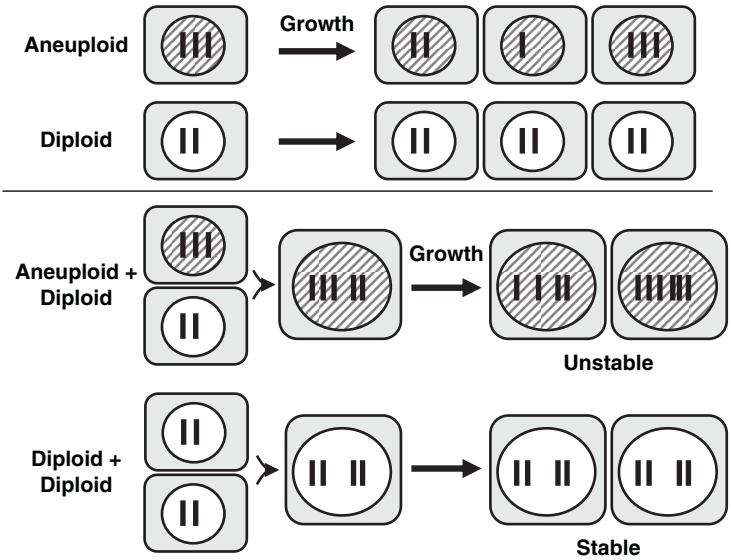

Fig. 4.3 Chromosomal instability in colorectal cancer cells. In vitro clonal expansion of an aneuploid cancer cell results in a cell line in which the individual cells have divergent numbers of chromosomes (upper panel). This instability defines the CIN phenotype. Diploid cell clones, in contrast, maintain a stable chromosome complement. Fusion of an aneuploid cell and a diploid cell results in a hybrid cell line with a large chromosome complement that exhibits CIN upon expansion (lower panel). This result demonstrates that CIN is dominant under these conditions. Fusion of two diploid cells similarly results in a hybrid cell with an abnormal number of chromosomes. Despite this abnormal complement, the progeny of this fused cell maintain numerical stability. This result shows that numerical abnormality does not, in itself, cause CIN

Chromosomes are lost when they fail to segregate equally to daughter cells during the process of mitosis. Chromosome gains occur when chromosomes are unevenly segregated and when they are aberrantly duplicated, suggesting that defects in the regulation of DNA replication might also contribute to CIN.

Aneuploid cancer cells most often have an excess of chromosomes. Cancer cells with a reduced number of chromosomes, which are sometimes termed *hypodiploid*, are relatively rare. The processes that underlie CIN do not appear to be biased; aneuploid cells have been shown to gain and lose chromosomes at equally high frequencies. Why then do aneuploid cancer cells most typically have a chromosome complement that is greater than the diploid number? The answer is probably related to cell survival. All chromosomes, with the exception of the Y chromosome, are essential. Loss of even one chromosome of a homologous pair can have lethal consequences, presumably due to the negative effects of reduced *gene dosage*. While CIN can cause the chromosome complement of a given cell to drop below the diploid number of 46, such a cell would be unlikely to survive and proliferate. Hypodiploid cell populations are therefore rare. The karyotypes of hypodiploid cancer cells typically reveal an extreme degree of structural rearrangement. Spectral karyotyping of such cells reveals individual chromosomes that contain material originating from multiple chromosomes. These derivative chromosomes can presumably maintain a vital gene dosage in the context of a reduced numerical complement.

Chromosome Instability Arises Early in Colorectal Tumorigenesis

Intriguing clues as to the role of aneuploidy during tumorigenesis have been provided by studies of colorectal tumors. CIN was first characterized in aneuploid cell lines that had been derived from established colorectal carcinomas, as described in the previous section and illustrated in Fig. 4.3. Subsequent studies have shown that even the smallest adenomas, less than 2 mm in size, have measurable allelic imbalances. These imbalances are a molecular indication of aneuploidy (which is a cytogenetic observation). Thus, evidence of aneuploidy can be seen in the earliest defined colorectal tumors.

In some tiny colorectal adenomas, allelic imbalances are evident in only a subset of the cell population. Notably, imbalances involving chromosome 5, which contains the locus for *APC*, are more likely to be present in every cell of a small tumor than are imbalances in other chromosomes. This observation is consistent with the preponderance of evidence that LOH involving the *APC* locus is the event that initiates colorectal tumorigenesis. Although the precise timing of CIN onset remains difficult to ascertain, the available data suggest that CIN occurs very early in the process of tumorigenesis, shortly after the biallelic loss of *APC*.

Does loss of *APC* lead directly to aneuploidy? Aberrant mitotic spindles have been detected in *APC*-null cells from experimental mice, suggesting that *APC* loss

may play a direct role in chromosome segregation. However, definitive evidence for such an affect during human colorectal tumorigenesis is currently lacking. It is also important to consider that, while aneuploidy is prevalent in most cancer types, *APC* mutations are mainly restricted to colorectal tumors. Thus, even if *APC* inactivation were found to be the proximal cause of CIN in colorectal cancers, loss of *APC* would clearly not be a general explanation for such a widespread phenomenon.

A general cause of aneuploidy would be expected to be present in many diverse cancer types. Genetic alterations of *P53* certainly fulfill this criterion. It has been suggested that *P53*, which is commonly mutated in a wide variety of cancers – including colorectal cancers – might play a critical role in maintaining chromosome stability. Evidence in favor of this hypothesis includes the overall prevalence of *P53* mutations, which approaches that of aneuploidy, and the finding that *P53* mutations appear to be more common in aneuploid cancers. However, there is also a significant body of evidence that calls a direct link between *P53* alteration and aneuploidy into question. There are many examples of chromosomally stable cancers with inactivated *P53* and, conversely, aneuploid cancers that have retained wild type *P53* alleles. Furthermore, the experimental mutation of *P53* alleles in chromosomally stable cancer cells does not cause these cells to express a CIN phenotype and become aneuploid. During colorectal tumorigenesis, the timing of *P53* inactivation (a late event) does not coincide with the onset of aneuploidy (an early event). The relationship between *P53* mutation and aneuploidy thus remains highly speculative in nature.

Chromosomal Instability Accelerates Clonal Evolution

The loss of genetic material can be lethal to a cell. There is clearly a lower limit to a chromosome complement, as attested by the relative paucity of hypodiploid cancer cells. There is also an apparent upper limit to how many chromosomes can be contained, maintained and transferred to progeny; few cancer cells have more than 90 chromosomes. Extreme levels of CIN would therefore be expected to be highly detrimental to the ongoing viability of a cell clone. Consistent with this prediction, the genes that are known to play central roles in mitosis and the mitotic spindle checkpoint have been found to be essential for viability.

A loss of genetic stability can clearly decrease cellular viability. However, a lower level of instability, such as that found in highly proliferative cancer cells, can augment clonal evolution and therefore increase the viability of cells in the changing environment of a growing tumor (see Fig. 4.4).

As described in Chapter 1, cancer cell clones evolve by the process of genetic mutation followed by successive waves of clonal expansion. How might CIN contribute to clonal evolution?

While many aspects of aneuploidy remain mysterious, one consequence of CIN is clear: CIN accelerates the late of LOH. As described in Chapter 3, the first step of tumor suppressor gene inactivation is the inactivation of one allele by mutation.

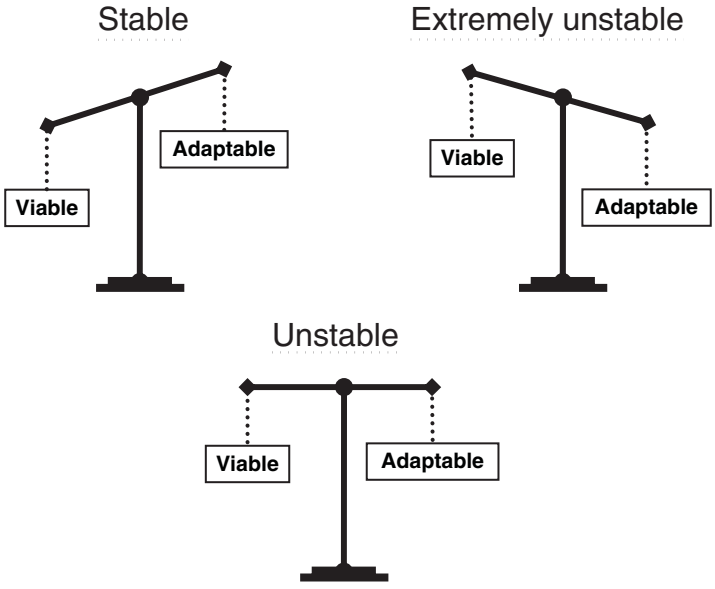

Fig. 4.4 Genetic instability balances viability against adaptability. Normal cells with a stable genome are highly viable, but not readily adaptable to changing environments. In contrast, hypothetical extreme levels of instability would promote adaptability but significantly impair viability. The level of instability in cancer cells appears to be optimal to facilitate adaptability and promote clonal evolution, while preserving an adequate level of viability to allow continued proliferation

The second step is the loss of the remaining wild type allele, known as loss of heterozygosity (LOH). LOH occurs either by an independent mutation, by mitotic recombination, or by loss of the chromosome that carries the remaining wild type allele. CIN would be predicted to directly increase the rate of the second step of tumor suppressor gene inactivation by increasing the rate of chromosome loss. In most cases, the loss of a chromosome that results in LOH is followed by duplication of the remaining homologous chromosome. The duplication process is also favored in cells with a CIN phenotype. Thus, the tendency of CIN cells to gain and lose chromosomes can contribute to two separate components of tumor suppressor gene inactivation: the accelerated loss of the wild type allele and the duplication of the mutant allele. The tendency to duplicate chromosomes inherent in the CIN phenotype might also contribute to the amplification of oncogenes.

In summary, the clonal evolution of cancer is punctuated by the progressive accumulation of genetic gains and losses. Genetic instability accelerates the rate of gain and loss and thereby promotes the progression of clonal evolution. The evolutionary advantage acquired by a cell clone that becomes genetically unstable is finely balanced against the disadvantages of instability. Too much instability is highly detrimental to cell viability. For example, the complete inactivation of most genes that contribute to mitosis, which would be predicted to cause an extreme level

of CIN, has been shown to be lethal. A more moderate level of CIN, which in many cancers has been measured as a loss or gain of 0.01 chromosomes per cell division, can accelerate the inactivation of tumor suppressor genes and the activation of oncogenes.

What Causes Aneuploidy?

The aberrant mitotic figures observed in aneuploid cell populations suggest that aneuploidy may result from intrinsic defects in the way that cancer cells divide. How might such defects arise? Current models are highly speculative and the ultimate answer to this question remains a topic of intensive investigation. Presently there are several possibilities that merit consideration:

Genetic alterations with direct effects on mitosis. The most obvious potential source of aneuploidy is alteration of the genes that control mitosis. Mutations in genes that encode proteins that participate in mitosis might be predicted to directly affect the ability of cells to maintain a stable number of chromosomes.

The segregation of chromosomes during mitosis is monitored by a mechanism known as the mitotic spindle checkpoint. In normal cells, this checkpoint functions to ensure that mitosis occurs in an orderly manner. Chromosomes must be properly aligned in metaphase cells and attached to the newly formed mitotic spindle, by a structure called the *kinetochore*, before chromosome separation can proceed. If one chromosome lags behind the others or fails to properly attach to the mitotic spindle, the mitotic spindle checkpoint becomes activated. This inhibitory pathway transiently blocks the subsequent steps of chromatid separation, thus allowing lagging chromosomes to 'catch up' and thereby be properly segregated.

Several genes that contribute to the mitotic spindle checkpoint have in fact been found to be mutated in cancers. The best-characterized examples are *hBUB1* and *hBUBR1*, both of which have been found to be mutated at a low frequency in several tumor types. Attempts to experimentally decrease the expression of these genes have successfully caused diploid cells to express a CIN phenotype. Germline mutations in *hBUBR1* have been found in individuals affected by *mosaic variegated aneuploidy*, a rare familial disease that causes inherent genetic instability and an increased risk of developing cancer. This rare disease provides conclusive proof that defined genetic changes can both cause aneuploidy and trigger the subsequent development of cancer.

Cancer-associated mutations have been found in genes that contribute to other aspects of mitosis as well. Mutations in genes that contribute to the kinetochore and to *centrosomes*, the organizers of the mitotic spindle, have been identified. While it has not yet been established whether kinetochore-associated mutations may actually cause CIN, homologs of these genes in model organisms such as yeast have demonstrated a role in the maintenance of chromosome stability.

While these cases conclusively demonstrate that single genetic alterations can induce CIN in an experimental setting, mutations mitosis-associated genes are not

found in the majority of cancers. The mutations that affect known regulators of mitosis are all rare. In the majority of aneuploid cancers, there is no established genetic alteration that would obviously cause a CIN phenotype.

A major focus of investigation in the study of aneuploidy has been the centrosome. Centrosomes contain the centrioles, the organizers of the mitotic spindle. Cancer cells are frequently observed to have abnormal numbers of centrosomes, or to contain centrosomes with structural abnormalities. These abnormalities have been found to correlate well with aneuploidy. Genetic factors that contribute to centrosome abnormalities in cancers, whether directly or indirectly, are poorly understood.

Genetic alterations with indirect effects on cell division. That the genetic basis for aneuploidy remains largely obscure may in part stem from a general lack of understanding of the many factors that contribute to and regulate mitosis and cell division. Studies of yeast have shown that mutations in over 100 genes can cause a CIN phenotype. Many of these genes had no previously appreciated link with cell division or chromosome stability. Thus, these important studies reveal that there is a great deal that remains to be learned about the genetic factors that dictate how cells grow and divide. It is possible that cancer-associated genetic mutations may affect chromosome stability in a manner that is not immediately obvious.

An example of such an unexpected relationship is the effect of alterations in *CCNE* and *CDK4* on chromosome stability. Both of these genes function to regulate the progression of the cell cycle, as will be described in Chapter 5. Amplification of *CCNE*, which encodes cyclin E, and inactivating mutations in *CDK4*, a cyclin-dependent kinase, are both found in cancers at low frequency. Introduction of these alterations in chromosomally stable cancer cells causes these cells to exhibit CIN. It had been understood for some time that Cyclin E and Cdk4 proteins function together as part of a multiprotein complex that regulates cell cycle transitions. More recently, evidence has emerged that that cyclin E and Cdk4 may regulate the mitotic spindle checkpoint, suggesting that the role of cyclin E and Cdk4 on chromosome stability might be more direct than was previously believed.

Another possible indirect mechanism for CIN involves the control of gene expression. It has been observed that the impairment of the mitotic checkpoint in cancers is frequently associated with changes in the levels of mitotic proteins. As tumor suppressor genes and oncogenes frequently regulate transcription, it is possible that mutations might indirectly affect mitosis by altering the expression of the proteins required for mitosis.

Inactivation of known tumor suppressor genes. Given the complexities of the maintenance of chromosome stability, it is quite possible that mutation of well-known tumor suppressor genes and proto-oncogenes might contribute to the development of aneuploidy. The studies of colorectal tumors described above underscore the many questions that remain. While aneuploidy has been temporally linked with *APC* inactivation in colorectal cancers, a clear role for *APC* in the stabilization of the chromosome complement has not been established. Similarly, the association of *P53* inactivation and aneuploidy is compelling, yet experimental disruption of *P53* has not supported a direct role.

Experimental evidence, largely derived from mouse models, suggests that inactivation of tumor suppressor genes involved in DNA recombination and repair might significantly contribute to aneuploidy. Examples of this type of gene include the breast cancer susceptibility genes *BRCA1* and *BRCA2*. Biallelic inactivation of *BRCA1* or *BRCA2* in mice leads to an increased incidence of tumors. These *BRCA1*- and *BRCA2*-null mouse tumors are highly aneuploid and exhibit centrosome abnormalities that are strikingly similar to those found in human cancers. It has not been elucidated whether the predisposing event, that is, *BRCA1* or *BRCA2* inactivation, or subsequent alterations that occur during the process of tumorigenesis are the proximal cause of aneuploidy. It is widely believed that defects in DNA repair might also contribute to the structurally aberrant chromosomes that are strongly associated with aneuploidy, and which are otherwise unexplained.

While studies of mouse tumor models have provided interesting links between known cancer genes and aneuploidy, the mechanism by which aneuploidy arises in the context of these alterations remains obscure. There remain more questions than clear answers. It appears that our understanding of cell growth and division as well as our understanding of cancer gene function are both limiting factors.

Given the current paucity of data to firmly support a genetic basis for aneuploidy it is worthwhile to consider several alternative nongenetic hypotheses:

Epigenetics. One idea is that epigenetic alterations to the genome (see Chapter 1) might play a central role in the stabilization of chromosomes. In this model, promoters of genes that contribute to the maintenance of genetic stability are silenced by cytosine methylation during tumorigenesis, thereby favoring the CIN phenotype and the development of aneuploidy. There is currently little evidence to either support or refute this idea. Consequently, a role for epigenetic changes in the CIN phenotype is largely supported as an alternative by a lack of genetic evidence.

Random aneuploidy. An older, but persistent hypothesis holds that aneuploidy is completely independent of genetic mutations or epigenetic changes. Contemporary proponents of this view, including Peter Duesberg, argue that aneuploidy arises as a random event that precedes genetic alterations. According to this model, the destabilizing effect of aneuploidy is sufficient to promote cellular evolution and ultimately cause all cancer phenotypes. The ‘random aneuploidy’ theory does not directly address the large and rapidly growing volume of mutational data that link specific cancer genes to cancers, nor does it explain how inherited mutations can markedly affect cancer predisposition.

Transition from Tetraploidy to Aneuploidy During Tumorigenesis

The development of CIN during tumorigenesis is one explanation of how cancer cells become aneuploid. There is a significant amount of data that suggest that other processes may also significantly contribute to aneuploidy.

A significant proportion of solid tumors are *polyploid*, that is, they have a chromosome complement that is a multiple of the haploid number. Most often, such cells have twice the diploid chromosome complement and are termed *tetraploid*. Tetraploidy would not be expected to result from gradual losses and gains of individual chromosomes. Instead, tetraploidy represents the aberrant duplication of the entire genome.

There are several lines of evidence that suggest that tetraploidy may represent an intermediate state during the development of some aneuploid cancers. Tetraploid cells are often seen in the lining of the esophagus in individuals prone to esophageal carcinoma. Esophageal carcinoma is known to evolve from a chronic inflammatory condition known as *Barrett's esophagus* by a series of histologically well-characterized steps. During the transition to cancer, the cells of the inflamed epithelium become first tetraploid and then aneuploid.

A high level of tetraploidization has also been observed in the colorectal mucosae of patients with *ulcerative colitis*, an inflammatory disease that strongly predisposes affected individuals to the development of colorectal cancer. Interestingly, the colorectal cancers associated with ulcerative colitis appear to arise from a precursor lesion that is morphologically distinct from a polyp, and that exhibits higher rates of *P53* inactivation and lower rates of *APC* and *K-RAS* mutations. Ulcerative colitis might therefore trigger a distinct sequence of mutations that define an alternative route to colorectal cancer. Interestingly, tetraploidization is seen in the context of ulcerative colitis but not in the polyps associated with FAP, nor in sporadic polyps. These data suggest that tetraploidization might contribute to the development of aneuploidy during the evolution of some tumors but not others.

The molecular basis of tetraploidization is incompletely understood, but appears to involve the failure of molecular mechanisms that link DNA replication with mitosis. Normal proliferating cells undergo mitosis after a single round of genomic DNA replication. Cells actively monitor this sequence of events. Cells with certain types of mutations are prone to replicate their genomes more than once without undergoing mitosis. Highly regulated cell cycle transitions are commonly referred to as *checkpoints*. Checkpoint regulators, which include *P53*, are frequently mutated in cancer cells. Thus, the uncoupling of DNA replication and mitosis by mutation of checkpoint regulators would be expected to increase the number of tetraploid cells. There is in fact experimental evidence that loss of *P53* can lead to an increase in the rate of tetraploidization. While *P53* inactivation has not been firmly established as a direct cause of aneuploidy, it may contribute to an intermediate stage of numerical aberration, in some cell types. The mechanisms by which cancer genes regulate checkpoints will be described in Chapter 5.

Some mitotic errors can alternatively lead to tetraploidy or to aneuploidy. Recent studies have demonstrated that chromosome missegregation during mitosis, which is often observed in aneuploid cells, sometimes leads to mitotic failure resulting in tetraploidization (see Fig. 4.5). Detailed studies of evolving breast tumors have suggested that aneuploidy is preceded by tetraploidy, and, furthermore, that tetraploidization is concurrent with the gradual loss and gain of individual chromosomes.

These observations suggest that, in some cancers, CIN and tetraploidization may both contribute to the development of aneuploidy.

In summary, a growing body of experimental evidence suggests that there are several pathways to aneuploidy, and that these pathways may be mechanistically interrelated. It is possible that multiple pathways may play a role in the development of every aneuploid cancer. Alternatively aneuploidy might evolve by a distinct pathway – or combination of pathways – in every evolving neoplasm, depending on the tissue of origin and/or the initiating mutation.

Multiple Forms of Genetic Instability in Cancer

Does genetic instability cause cancer or is it a merely a consequence of dysregulated cell growth? This is one of the oldest and most persistent questions in cancer genetics. In the case of aneuploidy, a causal role is suggested by its sheer prevalence

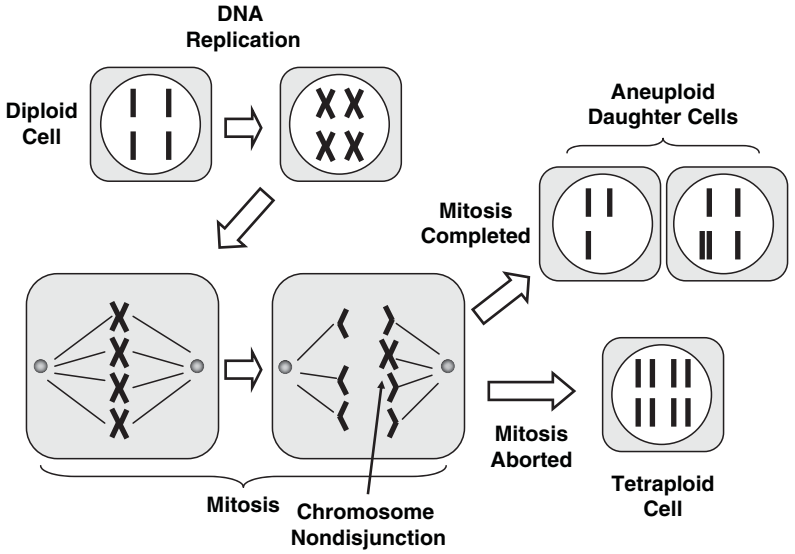

Fig. 4.5 Pathways to aneuploidy and tetraploidy. In a hypothetical model, chromosome nondisjunction can lead to aneuploidy or to tetraploidy. A diploid cell undergoes DNA replication prior to entering mitosis. (For illustrative purposes, only 4 chromosomes are shown.) Following the breakdown of the nuclear membrane, chromosomes align at the metaphase plate and attach to the mitotic spindle, which is organized by centrosomes. Sister chromatids separate during anaphase and migrate to opposite poles of the mitotic spindle; failure of this process results in chromosome nondisjunction. The activation of the mitotic spindle checkpoint by chromosome nondisjunction will cause mitosis to be delayed or aborted. Exit from mitosis then results in a tetraploid cell. Alternatively, failure of the mitotic spindle checkpoint allows mitosis to be completed. In this case, chromosome nondisjunction results in aneuploid daughter cells

in cancer and by the potential for CIN to accelerate the process of LOH. Additional evidence that aneuploidy actively participates in the evolution of cancers is provided, perhaps counterintuitively, by cancers that are not aneuploid.

While most solid tumors are composed of aneuploid cancer cells, the relatively small proportion of cancers that are not aneuploid exhibit defects in DNA repair. Every cell contains the machinery to repair DNA sequence errors that arise as a result of DNA polymerase errors or mutagen exposure. Defects in distinct DNA repair processes have been conclusively shown to significantly accelerate the development of several types of cancer. Summarized here, these repair processes and their inactivation in cancers will be discussed in detail in the sections that follow.

During DNA replication, most misincorporated bases are immediately corrected by the replicative DNA polymerase complex, which has a substantial, intrinsic proofreading capacity. As a result, the error rate of replicative DNA synthesis is estimated to be one in 10^{12} bases. This remarkable degree of fidelity implies that fewer than 1% of cells will acquire a single mispaired base during one complete S-phase. The rare misincorporated base that evades detection during DNA synthesis is processed by the mismatch repair (MMR) system. Approximately 15% of all colorectal cancers are estimated to have defects in MMR. All DNA repair systems, including MMR, involve the concerted activity of multiple proteins. Germline inactivating mutations in one of several MMR genes are the cause of *hereditary nonpolyposis colorectal cancer* (HNPCC), also known as Lynch syndrome. HNPCC is an autosomal dominant disease that, in addition to a highly elevated risk of colorectal cancer, also predisposes affected individuals to several additional types of epithelial cancers.

DNA replication errors represent an endogenous form of mutagenesis. In contrast, mutagens in the environment are an exogenous source of base changes. Altered bases that result from exposure to many types of environmental mutagens are processed by the nucleotide-excision repair (NER) system. Total inactivation of one of several NER genes causes a disease known as *xeroderma pigmentosum* (XP). XP, an autosomal recessive disease, strongly predisposes affected individuals to skin tumors in areas exposed to sunlight. In individuals homozygous for XP mutations, exposure to the UV component of sunlight causes unrepaired DNA base changes that would not occur in individuals with intact NER.

Defects in DNA repair processes such as MMR and NER cause genetic instability. Cancer cells with defects in MMR, for example, have a mutation rate that is between two and four orders of magnitude greater than that observed in normal cells with proficient MMR. Cellular defects in NER cause the accelerated accumulation of the UV signature mutations (see Chapter 1).

The changes to the genome that occur at high frequency in MMR- and NER-deficient cells are at the level of the DNA sequence. In contrast, the generation-to-generation changes in genome content associated with aneuploidy involve whole chromosomes or large chromosome segments that are visible upon karyotypic analysis. Despite these dissimilarities, both aneuploidy and DNA repair defects can accelerate the inactivation of tumor suppressor genes and the activation of oncogenes.

Analysis of familial cancers has provided critical insight into virtually every important aspect of cancer genetics. In the case of HNPCC and XP, these disorders have shown that genetic instability can exist in several forms and conclusively demonstrate that genetic instability directly promotes tumorigenesis. The cancers that occur in HNPCC and XP patients are clearly the result of the genetic instability caused by mutationally inactivated repair pathways.

It appears that genetic instability in some form is a universal feature of all cancers, both sporadic and inherited. Notably, the cancers associated with HNPCC and XP are rarely aneuploid. In general, aneuploidy and inactivated DNA repair pathways are mutually exclusive. As genetic instability clearly promotes tumorigenesis, aneuploidy is likely to be a causal factor in the majority of cancers in which it is observed.

Defects in Mismatch Repair Cause Hereditary Nonpolyposis Colorectal Cancer

The most common inherited form of colorectal cancer, and the most prevalent cancer predisposition syndrome known, is hereditary nonpolyposis colorectal cancer (HNPCC). HNPCC, also known as Lynch syndrome, is an autosomal dominant syndrome that is caused by inactivating germline mutations in the genes involved in the mismatch repair (MMR) system. Patients with HNPCC develop cancer at a young age, typically in the early to mid-forties but as early as the teens. Tumors in HNPCC patents occur disproportionately in the proximal segment of the colon. Although larger and less differentiated than the majority of colorectal tumors on average, HNPCC-associated colorectal cancers have a better outcome, as compared to stage-matched sporadic tumors. Carriers of germline HNPCC mutations are also susceptible to cancers in epithelial tissues of the uterus, small intestine, ovary, stomach, urinary tract, pancreas, biliary tract and brain.

HNPCC is a relatively common genetic disorder that was recognized as a distinct entity only recently. The delayed recognition of this syndrome occurred because colorectal cancer is very common in the general population, and because individuals affected by HNPCC do not have distinguishing traits other than an increased incidence of cancer. These factors contributed to difficulties of ascertainment.

Several families with numerous affected members were originally identified by the University of Michigan pathologist Aldred Warthin during the late nineteenth century. One family came to the attention of Warthin by way of his seamstress, who lamented that many of her relatives had died of cancer and predicted that she would likely die of cancer of the stomach, colon or uterus. This sad prophesy was realized when she died at a young age from endometrial carcinoma. Clusters of epithelial cancers in this family and others were documented and categorized in the 1960s and 1970s by Henry Lynch, for whom the syndrome was named. It was only in the 1980s that the concept of a familial cancer syndrome became widely accepted and studied.

The initial kindred identified by Warthin and subsequently analyzed by Lynch has exhibited an interesting shift in the types of cancers that develop in affected individuals. In the earlier generations of the family, gastric carcinomas were the predominant cancers that developed. Later generations increasingly developed colorectal carcinomas. This change in cancer incidence mirrors that which occurred in the general population over the same period. Presumably these changes are related to changes in the environment.

The search for the molecular basis of HNPCC involved complementary approaches employed by competing teams of researchers. In 1993, the discovery of a new and unusual DNA repair defect in colorectal cancer cells provided the critical clue. A group led by Manuel Perucho, while searching for genomic amplifications and deletions that might point to new oncogenes and tumor suppressor genes, instead found somatic alterations in the lengths of highly repetitive elements known as *microsatellites*. An independent group led by Stephen Thibodeau also came upon these altered microsatellite sequences and found that they were correlated with tumors of the proximal colon. This observation provided a potential connection between microsatellite abnormalities and HNPCC. Concurrently, a collaborative group led by Albert de la Chapelle and Bert Vogelstein was attempting to map the location of a tumor suppressor locus in Lynch kindreds using positional cloning methods. While mapping regions of LOH, the de la Chapelle/Vogelstein group also detected mutations in microsatellite sequences.

Microsatellites are repetitive DNA sequences widely distributed throughout the genome. Repeats are typically composed of between 10 and 100 units that are between one and four bases in length. The highly repetitive nature of microsatellites makes them unusually susceptible to mutation by slipped DNA strand mispairing (see Chapter 1). Mononucleotide repeats such as A_n or G_n and dinucleotide repeats such as $(CA/GT)_n$ are the most commonly affected by slippage, which causes either the expansion or the contraction of the number of bases within the repeat. The majority of mispaired bases are repaired by the proofreading mechanisms inherent to the replicative DNA polymerase complex. In normal cells, most of the mispaired bases that escape the proofreading process are subsequently resolved by the MMR system. The relatively high mutability of microsatellites renders them highly *polymorphic*. This attribute has made these repeat elements useful markers for a wide range of genetic analysis, including population studies and gene mapping.

Defects in MMR significantly impede the correction of mispaired bases and thereby increase the mutation rate. Microsatellite sequences in MMR-deficient cells are particularly susceptible to this effect and tend to expand and contract from generation to generation. This form of readily detectable hypermutability is known as *microsatellite instability* (MSI). MSI is a reflection of an increased mutation rate that affects the entire genome. The observation of MSI in colorectal cancer cells illuminated an entirely new mechanism of tumorigenesis. MSI is not restricted to colorectal tumors but can be detected in extracolonic tumors, such as gastic, endometrial, and other cancers that occur in HNPCC patients.

Interestingly, the pivotal insights into the genetic basis for MSI were provided not by studies of cancers, but by studies of model microorganisms. MSI was found

to strongly resemble mutation patterns previously found in bacteria and yeast that were defective for MMR. In the bacterium *E. coli*, the MMR system is known as the MutHLS pathway. This system functions to recognize mismatched bases that arise during DNA replication, to excise the mismatched and neighboring bases and to trigger the resynthesis of a defined region, or 'patch', of DNA. This pathway is dependent upon several genes, including *MutS* and *MutL*. Biochemical studies demonstrated that dimeric MutS protein detects the mismatch and recruits a MutL dimer to the repair site.

Eukaryotic homologs to bacterial *MutS*, designated *MutS homolog* or *MSH*, were found in yeast (*yMSH*) and in human cells (*hMSH*). In yeast, mutation of *MSH* genes was found to lead to 100- to 700-fold increases in the mutation rate of dinucleotide repeats. The revelation that MSI was related to MMR defects provided a critical clue as to the identities of the HNPCC genes. Shortly following the discovery of MSI in colorectal cancers, groups led by Richard Kolodner and Bert Vogelstein identified a human *MutS* homolog, *hMSH2*, on chromosome 2. Germline mutations of *hMSH2* were found in a substantial proportion of Lynch kindreds. Additional MMR genes were similarly identified by positional cloning and by virtue of known interspecies protein and DNA sequence homologies.

MMR is a basic biological process that is evolutionarily conserved. Human cells contain at least five *MutS* and four *MutL* homologs. Five of these genes have been shown to play a role in MMR and to cause HNPCC when mutated (see Table 4.1). While the first steps of MMR in bacteria involve the activity of MutS and MutL homodimers, the human proteins form heterodimers in various combinations. The different specificities of these complexes allow the recognition of different substrates (see Fig. 4.6). *hMSH2* plays a fundamental role in the recognition and binding of mispaired bases, while *hMSH3* and *hMSH6* appear to modify the specificity of this recognition. The MutL homolog *hMLH1*, which is recruited to the repair site by the MutS homologs, functions as molecular matchmaker. As a hetrodimeric complex with *hPMS2, hMLH1* couples mismatch recognition with downstream steps of MMR, which include 'long patch' regional DNA excision, repair synthesis and religation. The role of PMS1 in this process remains to be determined.

Genetic analysis of the human MMR genes revealed that mutations in *hMSH2* and the *MutL* homolog *hMLH1* account for the majority of documented Lynch syndrome mutations. As has been shown to be the case with other familial cancer syndromes such as familial breast cancer, HNPCC kindreds with different mutations exhibit distinct patterns of disease. *hMSH2* mutations are more strongly

Table 4.1 MMR genes involved in HNPCC

E. coli gene	*H. sapiens* homolog	Chromosomal location	Mutated in HNPCC (%)	Predisposition
MutS	*hMSH2*	2p21–22	40	Typical HNPCC
	hMSH6	2p16	10	Atypical HNPCC
MutL	*hMLH1*	3p21	50	Typical HNPCC
	hPMS1	2q31–33	Rare	Typical HNPCC
	hPMS2	7p22	<2	Turcot Syn

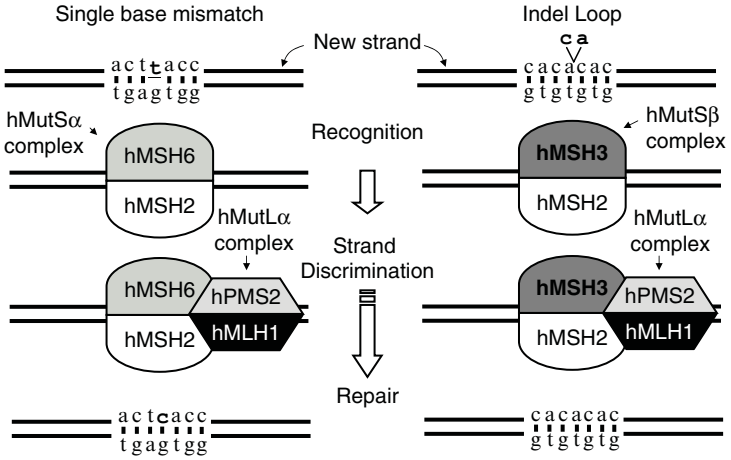

Fig. 4.6 Human DNA mismatch repair. Mispaired bases (shown in bold) create a physical deformity in the DNA double helix which is recognized by MutS homologs. Single base mismatches are recognized by hMSH2-hMSH6 heterodimers, which are known as the hMutSα complex (right). Looped-out bases caused by short insertions and deletions (indels) cause a distinct structure primarily recognized by hMSH2-hMSH3 dimers, known as the hMutSβ complex (left). Note that there is some overlap in recognition affinities of hMutSα and hMutSβ. The MutLα complex composed of hMLH1 and hPMS2 is recruited to the repair site and functions to determine which DNA strand contains the error. The MutLα complex communicates this information to downstream repair proteins that excise a region of DNA known as a long patch, resynthesize the damaged strand and ligate repaired strands. Many details of this process remain to be fully understood; it is likely that additional components remain to be discovered

associated with extracolonic cancers than are mutations in *hMLH1*. Mutations in the *MutS* homolog *hMSH6* are found in a minority of kindreds. Germline mutation of *hMSH6* is associated with an atypical form of HNPCC, characterized by a somewhat older mean age of cancer diagnosis and a risk of endometrial cancer that is higher than that conferred by mutations in *hMSH2* or *hMLH1*. A small number of germline mutations have been found in the *MutL* homolog *hPMS2*. These mutations are associated with Turcot syndrome, an HNPCC syndrome-related disorder characterized by an increased risk of brain and early-onset colorectal cancers. The biological basis of gene-specific differences in syndrome phenotypes is not well understood.

While HNPCC was recognized as a disease entity by the identification of large kindreds with multiple affected individuals, many additional carriers of HNPCC gene mutations have been identified by population based genetic screening. Genetic testing has revealed that MMR gene mutations are highly prevalent among individuals with inherited colorectal or endometrial cancers, most of whom are not known members of HNPCC kindreds. Among cancer patients age 50 and younger that have at least one relative with cancer, nearly one quarter have germline mutations in one of the two highly mutated HNPCC genes, *hMSH2* or *hMLH1* (see Table 4.2).

Table 4.2 Prevalence of MLH1 and MSH2 mutations among cancer patients. (Data from Myriad Genetic Laboratories, October 2005.)

	Family history	
Personal history	No affected relatives (%)	≥ 1 Relative affected
Colorectal cancer (age < 50)	9.1	22
Colorectal cancer (age ≥ 50)	<5	15
Endometrial cancer (age < 50)	11	23
Other HNPCC-associated cancer	6.5	16

The HNPCC genes are tumor suppressor genes. In terms of their stepwise inactivation, they share the general characteristics of the tumor suppressor genes discussed in Chapter 3. HNPCC genes are present in the germline of cancer-prone families as a single mutated allele. The wild type allele is lost somatically, giving rise to the cellular MMR defect.

Unlike the gatekeeper defects caused by biallelic inactivation of *APC* or *RB*, the inactivation of HNPCC genes does not directly affect cell proliferation. Rather, HNPCC gene inactivation causes defective DNA repair, a form of genetic instability. Cells with MMR defects have a higher mutation rate than normal cells and thus have an increased probability of acquiring tumor suppressor gene and proto-oncogene mutations. Mutations in HNPCC genes, like those in other tumor suppressor genes that are involved in DNA repair, cause caretaker defects.

The nature of the cellular defect in HNPCC is clearly reflected in the pathogenesis of the disease (see Fig. 4.7). HNPCC is readily distinguished from the less common familial adenomatous polyposis (FAP) by tumor number. As described in Chapter 3, FAP is characterized by a very large number of polyps. Although each polyp has only a small chance of developing into an invasive cancer, the cumulative risk caused by the large number of polyps in FAP-affected individuals makes the development of cancer all but inevitable. In contrast, individuals affected by HNPCC do not exhibit polyposis; patients with HNPCC develop polyps at approximately the same rate as the general population. However, the underlying genetic instability greatly increases the chance that a given polyp will progress to a cancer.

The classical tumor suppressor genes, exemplified by *APC, RB* and *P53*, are mutated in the germline of cancer-prone kindreds, but are also mutated somatically in sporadic cancers. The MMR genes are similarly involved in both inherited and sporadic cancers. In unselected samples, MSI occurs in colorectal cancer at a rate of approximately 13% . In the majority of MSI-positive cases, germline mutations in MMR genes are not detected in normal tissues from the same patient. Two important conclusions can be drawn from these observations: (1) the majority of MSI-positive cancers are not associated with HNPCC, a heritable disorder, and (2) a sizable proportion of all sporadic cancers acquire mutations in MMR genes.

The high rate of colorectal cancer in the general population, as well as technical challenges inherent to the comprehensive analysis of the five different genes

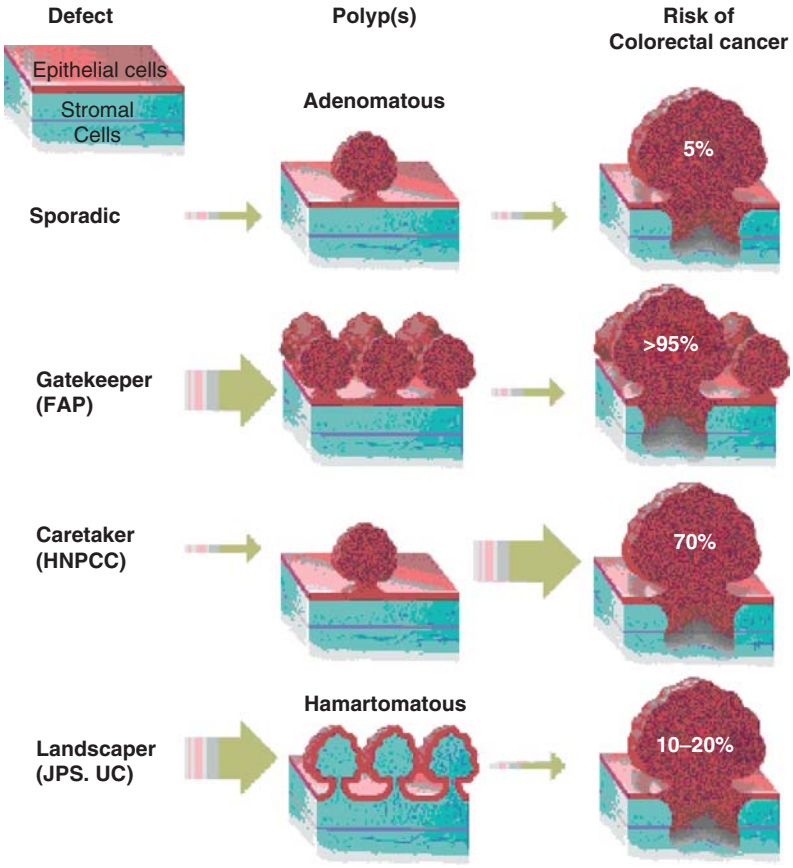

Fig. 4.7 Pathways to colorectal cancer. Colorectal cancer can be caused by genetic defects that disrupt 'gatekeeper', 'caretaker', or 'landscaper' processes in the colon. Adenomatous polyps develop in half of the general population by age 70. A minor fraction of these benign tumors develop into invasive cancers, yielding a lifetime cancer risk of 5%. Patients with familial adenomatous polyposis (FAP; Chapter 3) have an inherited defect in *APC*, which causes the development of 10^2–10^3 polyps at a young age. Each polyp has a low malignant potential, but the large number of polyps greatly increases the overall risk of cancer. In contrast, patients with HNPCC develop polyps at a rate similar to that of the general population. The increased mutation rate caused by defective MMR causes an increased proportion of these benign tumors to develop into cancers. Landscaper defects caused by juvenile polyposis syndrome (JPS; Chaper 3) or ulcerative colitis (UC; Chapter 1) create an abnormal microenvironment that increases the probability that neighboring epithelia will become neoplastic. Bold arrows indicate the step accelerated in each class of tumor. (Reprinted with permission from Kinzler and Vogelstein *Science* **280**, 1036. Illustration by K. Sutliff. Copyright 1998 AAAS.)

associated with defective MMR, has made it difficult to accurately determine the prevalence of HNPCC. In addition to mutations in MMR genes, most commonly in *hMSH2* and *hMLH1*, MMR defects have also been found to be caused by epigenetic

silencing of *hMLH1* in some cases. By the current best estimates, approximately 3% of colorectal cancers in the United States are attributable to hereditary defects in MMR, which cause HNPCC. The remaining 10% of MSI-defective colorectal cancers are caused by somatically acquired MMR defects. It remains a possibility that additional MMR genes, yet to be identified, are mutated in some of the cancer prone kindreds in whom MMR gene mutations have not been found.

An understanding of the mutational mechanisms of MMR and the genetic basis of HNPCC are likely to provide direct benefits for cancer-prone individuals. HNPCC is a common genetic disorder that causes a significant number of cancers. Analyzing tumor samples for the presence of MSI can aid the discovery of potential HNPCC kindreds, while searching for germline HNPCC mutations is a strategy for identifying HNPCC-affected individuals before they develop invasive cancers. The use of genetic analysis to uncover HNPCC promises to have a significant impact on public health.

From a theoretical perspective, HNPCC reveals an unambiguous role for the loss of genetic stability in the process of tumorigenesis. Genetic instability caused by MMR-deficiency accelerates the rate at which tumors develop into cancers. This explains why individuals affected by HNPCC develop cancer at a younger age.

Mismatch Repair-Deficient Cancers Have a Distinct Spectrum of Mutations

MMR defects cause an increased mutation rate. It can be inferred that mutation of MMR genes leads to increases in the rates at which tumor suppressor genes are inactivated and proto-oncogene are activated. Indeed, the specific mutations found in MSI-positive/MMR-deficient cancers reflect a unique mechanism of mutagenesis.

It appears that the same molecular pathways are disrupted by mutation in MSI-positive/MMR-deficient and MMR-proficient tumors alike. For example, mutations in *APC* initiate the process of colorectal tumorigenesis in both tumor types. However, MSI-positive/MMR-deficient cancers have a higher rate of frameshift mutations at intragenic mononucleotide tracts, most strikingly at A_n (see Table 4.3).

Table 4.3 Representative inactivating *APC* mutations found in MSI-positive/MMR-deficient colorectal tumors. Insertions and deletions are indicated in bold type. Mutations frequently occur in sequences that contain mononucleotide tracts. (Data from Huang et al. (1996) *PNAS* **93**, 9049–9054.)

Family history	*APC* Codon	Nucleotide change	Target sequence
Sporadic	758	1 bp deletion	AAC**a**AAAAGCC
Sporadic	773	2 bp deletion	GAAAC**tt**TTGAC
Sporadic	801	1 bp deletion	TATG**t**TTTTGAC
HNPCC	847	1 bp insertion	TCTG**(A)**AAAAAGAT
HNPCC	907	1 bp deletion	TCT**g**GGTCT
HNPCC	941	1 bp insertion	TCGG**(A)**AAAATTCA
Unknown	975	1 bp deletion	GGT**a**AAAGAGGT

The overall prevalence of detectable *APC* mutations is also somewhat lower in MSI-positive/MMR-deficient tumors. In MSI-positive/MMR-deficient colorectal cancers with wild type *APC*, a mutation in the *CTNNB1* gene, which encodes β-catenin, has been found to phenocopy the loss of *APC*. This observation provides one explanation for the lower *APC* mutation rate in MSI-positive/MMR-deficient tumors. A similar situation has been found to affect the pathway mediated by *K-RAS*. The oncogene *K-RAS* is commonly mutated in MMR-proficient colorectal cancers, but mutated less often in MSI-positive/MMR-deficient cancers. Such cancers often contain mutations of a gene known as *BRAF*, which encodes a protein that functions downstream of K-RAS.

The genetic alteration most frequently associated with MMR-deficiency is mutation of the gene *TGFBR2*, which encodes the transforming growth factor β type II receptor (TGFβ-RII). The coding region of *TGFBR2* contains an A_8 tract that is mutated in 85–90% of MSI-positive tumors. The TGFβ-RII protein is a cell surface receptor functions upstream of the tumor suppressor gene *SMAD4*. These two proteins are critical components of a molecular pathway that communicates growth-inhibitory signals from the cell surface to the nucleus, where gene expression is regulated. The *TGFBR2* mutations in MSI-positive cancers have a similar functional result as the *SMAD4* and *SMAD2* mutations that occur in MMR-proficient cancers.

P53 is mutated at a lower frequency in MSI-positive/MMR-deficient cancers. One possible explanation for this observation is that an alternate gene in the *p53* pathway may be preferentially mutated in cells with MMR defects. One candidate is *BAX*, a gene involved in the process of programmed cell death, or apoptosis. The coding region of *BAX* contains a G_8 tract that is mutated in approximately 50% of MSI-positive/MMR-deficient cancers. *BAX* has been found to be transcriptionally transactivated by *p53* in some cell types. It may be possible that mutation of *BAX* in some tissues eliminates the selective advantage of *P53* mutation. However, it should be noted that *BAX* mutations appear to cause only a small subset of the cellular phenotypes that are associated with *P53* inactivation.

That a genetically defined pathway can be alternately inactivated by different gene mutations is a central principle of cancer genetics. The molecular mechanisms of signal transduction and the roles of proto-oncogenes and tumor suppressor genes in signaling pathways will be discussed in detail in Chapter 5.

Defects in Nucleotide Excision Repair Cause Xeroderma Pigmentosum

Nucleotide-excision repair (NER) is a versatile type of DNA repair required for the processing of a variety of base lesions caused by environmental mutagens. The most important mutagen that affects individuals with defects in the NER system is the ultraviolet (UV) component of sunlight. As described in Chapter 1, UV light causes

a number of distinctive DNA lesions, including cyclobutane pyrimidine dimers and (6–4) photoproducts. These damaged bases are normally removed from the genome by the process of NER. Bulky pyrimidine and purine adducts formed by psoralen derivatives, the chemotherapeutic drug cisplatin, and the polycyclic carcinogens such as acetylaminofluorene are also processed by the NER system.

Xeroderma pigmentosum (XP) is a rare disease caused by defective NER. First described by Moriz Kaposi as a skin disorder in 1870, XP was recognized to be often associated with neurological abnormalities by Albert Neisser in 1883. Interestingly, Neisser was also the discoverer of the etiological agent of gonorrhea, the bacterium *Neisseria*.

XP is characterized by sensitivity to sunlight, a symptom known as *poikilodermia*, and the development of skin neoplasia. Symptoms most commonly present at the age of 1–2 years, concomitant with the first exposures to sunlight, and include severe sunburns and subsequent freckling. The accumulation of unrepaired DNA lesions results in the progressive degeneration and atrophy of sun-exposed skin and eyes. *Xeroderma* literally refers to the typical parchment-like appearance of exposed skin in affected individuals, while *pigmentosum* describes the pigmentary abnormalities commonly found in these patients (see Fig. 4.8). XP patients are prone to cataracts and both benign and malignant ocular tumors. Approximately 20–30% of XP patients develop a variety of neurologic abnormalities that are often progressive.

Fig. 4.8 Xeroderma pigmentosum. Sun-exposed areas of skin typically have a prematurely aged appearance. Multiple scars and lesions mark sites of treated and developing carcinomas. (Reprinted from the Atlas Genet Cytogenet Oncol Haematol, October 2000, Viguie, C. Xeroderma pigmentosum. Image by Daniel Wallach, with permission of the Atlas.)

XP patients develop an array of benign lesions that arise from various cell types present in the skin. These include the keratinocytes and fibroblasts that are the main structural components of the skin, but also cells of the cutaneous vasculature and adipose tissue. Some, but not all, of these lesions are premalignant and thus have the potential to develop into cancers. XP patients most commonly develop basal and squamous cell carcinomas, but also have a significantly increased risk of developing melanoma, a deadly form of skin cancer that develops from the pigmented melanocytes (see Chapter 6). The onset of cancer in XP patients is strikingly accelerated. Skin cancers occur at a median age of 8 years, which is 50 years earlier than in the general population. The overall risk of skin cancer before the age of 20 is increased more than 1,000-fold in XP patients. Many XP patients die of cancer; only 70% survive past the age of 40.

XP patients also have a moderately increased propensity to develop various solid tumors, most commonly brain cancers. The development of these internal tumors, as well as the neurologic abnormalies that are often detected in XP patients, indicate the complex involvement of NER genes in processes unrelated to the resolution of UV lesions. Chemical mutagens are thought to play a significant role in the etiology of internal cancers that arise in XP patients.

XP occurs at a frequency of approximately 1 in 1,000,000 in the USA and Europe, while tenfold to 100-fold higher frequencies are observed in Japan and North Africa. Unlike the other cancer predisposition syndromes discussed in preceding sections, XP exhibits an autosomal recessive mode of inheritance. Patients are thus homozygous for the primary genetic mutations that underlie the disease. *Consanguinity* between patients' parents has been reported in a significant proportion of cases.

Every cell in a typical XP patient has defective NER. Cellular defects in NER are readily detectable in the laboratory. These two features greatly aided the discovery of the genetic defects that underlie XP. When cultured fibroblasts from normal individuals are exposed to UV light, the damaged DNA triggers DNA synthesis associated with lesion repair, also known as *unscheduled DNA synthesis* (UDS). The activation of NER can therefore be directly detected by measuring the uptake of the DNA precursor [^{3}H]-thymidine (see Fig. 4.9). Fibroblasts from XP individuals are defective for NER and therefore do not exhibit UDS.

A significant subset of XP patients exhibit normal levels of UDS and thus have no obvious defects in NER. This variant form of XP are nonetheless characterized by an increased rate of mutagenesis upon UV exposure. Designated XP-V, this group is clinically indistinguishable from other XP patients. Further analysis of XP-V cells has suggested that at least some patients in this catagory have a repair defect that is manifest only during DNA replication. The XP-V patient group exhibits a range of disease severity, with a general lack of neurological abnormalities. The substantial clinical heterogeneity of XP-V suggests that this group might be similarly heterogeneous from a genetic perspective.

Employed in seminal experiments throughout the history of cancer genetics, the technique of cell fusion has been a powerful tool for understanding the nature of cancer cell defects. This strategy proved enormously successful in categorizing the

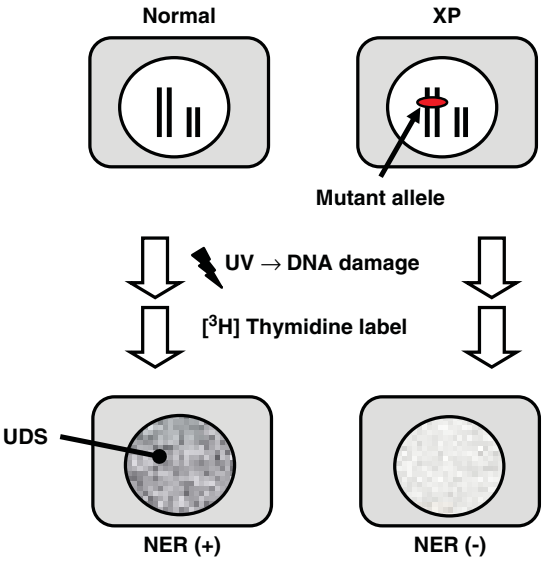

Fig. 4.9 DNA damage triggers unscheduled DNA synthesis. When challenged with UV light, cells harboring normal XP alleles synthesize DNA during DNA repair. This non-replicative synthesis is known as unscheduled DNA synthesis (UDS). Normal cells will incorporate radio-labeled thymidine at repair sites, which can then be visualized by autoradiography. Cells from XP patients contain biallelic mutations (shown in red) in a gene involved in the NER process. XP cells do not exhibit UDS because they are NER (-). For simplicity, only two homologous pairs of chromosomes are shown

gene mutations that cause XP (see Fig. 4.10). Among unrelated XP patients, cells from one patient will functionally complement cells from another patient upon fusion only if the mutated genes in each patient are different. By this technique, XP patients could be categorized into a total of seven distinct complementation groups, designated XP-A through XP-G. (The XP-V group has no UDS defect and cannot, by definition, be complemented.) Each complementation group is believed to be defined by a single, distinct mutated gene.

The finding that defective NER can be restored by complementation facilitated the cloning of genes that underlie XP. New reagents were developed for this effort, including mutant rodent cell lines that had defects in DNA repair and were UV-sensitive. Rodent cells could be transformed with either human genomic DNA or with human DNA libraries, allowing the recovery of UV-resistant clones. In a period spanning the mid-1980s and the mid-1990s, groups led by Larry Thompson, Dirk Bootsma and Jan Hoeijmakers cloned human genes that would complement the NER defects of the different rodent cell lines. These genes were designated human *excision repair cross-complementing*, or *ERCC* genes. Several of these were demonstrated to be mutated in XP complementation groups. Randy Legerski and colleagues were able to directly complement a human cell line from an XP-C

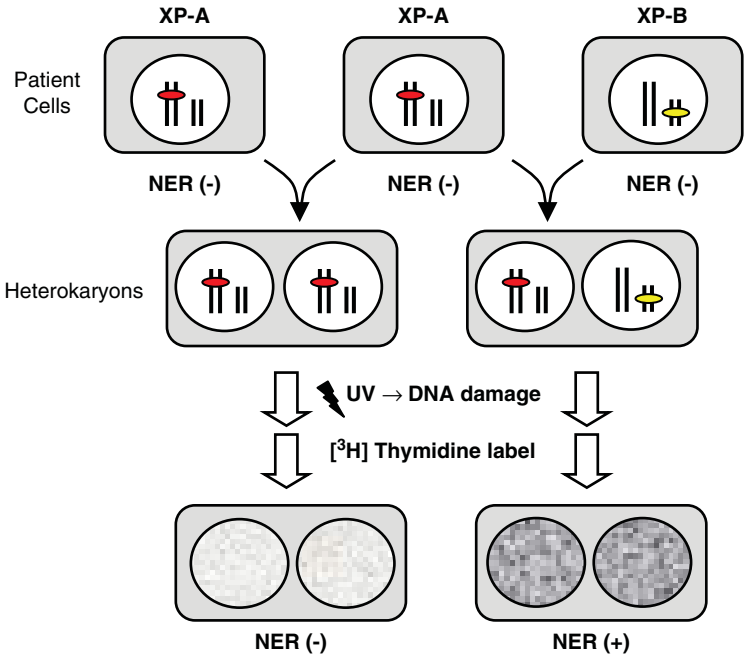

Fig. 4.10 Cellular complementation reveals the genetic heterogeneity of XP. Cells from three individual XP patients are shown (top). Pairs of cells are fused, creating hybrid cells. Some hybrids, known as heterokaryons, contain nuclei from both patients. The fusion of cells that contain the same XP gene mutation (shown in red) results in hybrids that remain NER(-), as indicated by the absence of UDS. When cells that contain two different mutations (red and yellow) fuse, each nucleus contributes complementary wild type genes to the heterokaryon. NER is thus restored in these hybrids. Such experiments allow the categorization of XP patients into distinct complementation groups, which in this example are denoted XP-A and XP-B. For simplicity, only two homologous pairs of chromosomes are shown

patient with a gene that was initially designated *XP-C complementing clone*, or *XPCC*. The gene that complements the XP-A group was isolated by a group led by Yoshio Okada. The last complementation group to be genetically characterized, largely by the efforts Stuart Linn and his coworkers, was XP-E. Though there appears to be some heterogeneity within this group, it appears that the majority of XP-E patients have mutations in *DDB2*, which encodes a DNA damage-specific DNA binding protein. Each gene conclusively demonstrated to be mutated in a specific complementation group has been subsequently so designated (see Table 4.4). For example, the gene *ERCC3* has been found to functionally complement the patient group XP-B, and is therefore usually referred to as *XPB*.

The extensive biochemical characterization of XP-V cells conducted independently by the laboratories of Fumio Hanaoka and Louise Prakash revealed a functional defect in DNA polymerase η. This replicative DNA polymerase,

Table 4.4 The XP complementation groups and genes

Comp. group	Relative UDS	Gene	Location	Mutated in XP	Function of encoded protein
XP-A	<2%	*XPA*	9q34	25%	Binds UV-damaged DNA, forms demarcation complex
XP-B	3–7%	*XPB (ERCC3)*	2q21	Rare	3′ → 5′ DNA helicase, TFIIH subunit
XP-C	10–25%	*XPC*	3p25	25%	Binds UV-damaged DNA, initates GGR
XP-D	25–55%	*XPD (ERCC2)*	19q13	15%	5′ → 3′ DNA helicase, TFIIH subunit
XP-E	50%	*XPE (DDB2)*	11p12	Rare	Binds UV-damaged DNA during recognition step
XP-F	15–30%	*XPF (ERCC4)*	16p13	6%	Endonuclease, makes 5′ incision
XP-G	10%	*XPG (ERCC5)*	13q32	6%	Endonuclease, makes 3′ incision
XP-V	100%	*XPV (POLH)*	6p21	21%	Translesion DNA synthesis during subsequent S-phase

encoded by *POLH*, is involved in translesion DNA synthesis during S-phase. The role of DNA polymerase η in UDS-proficient XP was confirmed by the finding of *POLH* mutations in XP-V patients.

NER is a multistep process that involves the concerted function of numerous protein complexes. There are two distinct modes of NER depending on whether a lesion is located within an expressed gene. DNA lesions occurring within actively transcribed regions of the genome, which are presumably more detrimental to cellular viability, are repaired by a process known as *transcription-coupled Repair* (TCR). In the remainder of the genome, lesions are removed by an NER sub-process known as *global genome repair (*GGR). Some specific types of lesions are more efficiently removed from transcribed regions, via TCR, than they are from the rest of the genome via GGR. In the majority of XP cases, the genetic defect affects both TCR and GGR. The exception is the XP-C group, in which only GGR appears to be affected. That *XPC*-mutant individuals are highly prone to cancer reveals the importance of GGR in tumor suppression.

The first step in NER is the recognition of damaged bases (see Fig. 4.11). In TCR, the initial detection of a lesion occurs when the RNA polymerase II holoenzyme complex stalls while transcribing the affected DNA strand. This transcription complex is then rapidly displaced in favor of repair proteins. In GGR, DNA damage recognition is accomplished by a distinct NER protein complex containing the XPC protein. The protein encoded by *XPE* plays a role in DNA damage recognition by both pathways.

Although the mechanism of recognition of damaged bases differs in transcribed genes and untranscribed regions of the genome, the subsequent steps of NER are

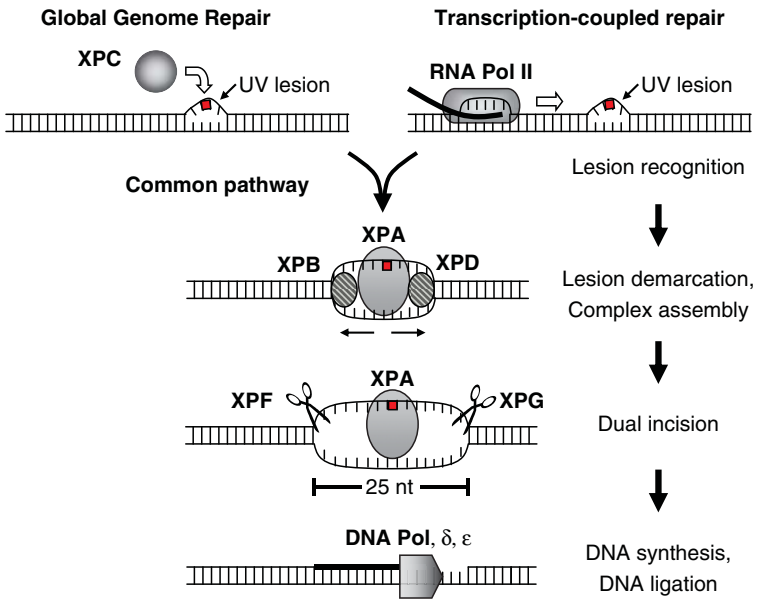

Fig. 4.11 The NER process. Shown are the roles of the XP genes in the four stages of NER. A UV-induced lesion (red) is recognized either by XPC (GGR pathway) or by the RNA Pol II complex. (TCR pathway). XPE (not shown) plays a role at the initiation stage of both pathways. The DNA around the region is opened by the TFIIH helicases XPB and XPD, and demarcated by XPA. The damaged strand is then incised by the endonucleases XPF and XPG, allowing the release of the damage-containing oligonucleotide. Finally, the double helix is restored by the sequential activity of DNA polymerases δ and ε, and a DNA ligase. Many additional repair and replication proteins, which are not known to be defective in XP patients, are also involved in NER. See text for additional details

common to both pathways. Once a site of DNA damage is recognized, the adjoining paired bases are separated by the helicase activities of XPB and XPD, and the site is demarcated by the binding of the XPA protein. Both XPB and XPD are known components of an evolutionarily conserved transcription factor complex known as Transcription Factor II H (TFIIH). TFIIH is a protein complex that controls the initiation of basal gene transcription. The shared requirement of XPB and XPD for these processes reveals the close relationship between NER and the basic mechanisms of gene expression.

The region opened by the TFIIH helicases, spanning about 25 bases, is accessible to the endonucleases XPF and XPG. These enzymes cut the DNA backbone at the junction created by the helicases and thereby excise the bases on the damaged strand. The gap left by the incision process is then filled in by DNA polymerases that specifically function in DNA repair. The newly synthesized DNA strand is covalently joined to the double helix by a DNA ligase.

NER Syndromes: Clinical Heterogeneity and Pleiotropy

XP is a heterogeneous disorder, from both a clinical and a genetic standpoint. Much of this overall heterogeneity is due to the differences between the complementation groups. Patients in the XP-C group are affected with what is often termed the 'classic' form of XP. Disease caused by *XPC* mutation is restricted to the skin and eyes, and is dependent on sun exposure. Mutations in *XPA* cause a more severe form of the disease, which is manifest from birth and features progressive neurologic degeneration. A small number of individuals exhibit severe neurological disease along with dwarfism and immature sexual development. This form of XP, generally found within the XP-A group, has been termed the *DeSanctis–Cacchione syndrome*.

While there are significant differences between XP-A and XP-C, the presentation of disease *within* each group is fairly uniform. In contrast, several of the XP complementation groups exhibit a significant degree of heterogeneity associated with a single gene. For example, patients in the XP-D group show a widely diverse spectrum of disease severity. Some patients with *XPD* mutations are clinically indistinguishable from those in the XP-A group, while others have a more mild form of the disease resembling classic XP, as seen in the XP-C group. These findings demonstrate that different mutations in the same gene can cause distinct phenotypes. Similarly complex genotype/phenotype relationships are seen in the cancers caused by mutations in other tumor suppressor genes, such as *APC* (see Chapter 3).

Mutations in NER genes can also cause two related syndromes that clinically overlap with XP. Unlike XP, these rare disorders are not typically associated with an increased risk of cancer:

Cockayne syndrome (CS). Patients with CS typically exhibit sun sensitivity, short stature, severe neurological abnormalities, dental caries, cataracts and a wizened appearance. The average life span of CS patients is 12 years. Cells from CS patients exhibit normal UDS. CS is caused by mutations in either *CSA* or *CSB*, which are involved in TCR. Accordingly, CS cells exhibit defects in TCR, but have normal GGR. A small number of patients with the characteristics of CS have an associated cancer predisposition; this syndrome has been referred to as XP-CS complex and is caused by mutations in *XPB* and *XPD*.

Trichothiodystrophy (TTD). The clinical features of TTD include most of those that define CS. In addition, TTD patients have brittle hair and dystrophic nails that are caused by a reduced content of cysteine-rich, sulfur-containing matrix proteins. TTD is a heterogeneous disorder and at least seven previously described disorders resemble or are identical to TTD. TTD cells exhibit impaired UDS, and are defective in both TCR and GGR. TTD can be caused by some mutations in *XPB* or *XPD*, or by mutations in *TBF5*, which encodes another subunit of the TFIIH transcription complex.

XP, CS and TTD reveal the diverse phenotypic manifestations of NER gene mutation. Mutations in some NER genes are clearly pleiotropic. Distinct mutations

in *XPB* and *XPD* have been shown alternatively cause XP or XP-CS or TTD, which collectively encompass a broad range of disease phenotypes. Many of these phenotypes are unrelated to cancer.

The basis of NER gene pleiotropy is thought to reside in the overlapping activities of the TFIIH complex, which functions in both NER and basal transcription. All of the genes that encode TFIIH are essential for viability, and are therefore never found to be completely inactivated by mutation. The mutations that are found in *XPB* and *XPD* are not inactivating mutations, but rather have subtle effects on TFIIH function. Different mutations in *XPB* and *XPD* can separately affect the two TFIIH-related activities and thus cause distinct diseases. The mutations that underlie CS and TTD are thought to predominantly cause the defective function of the TFIIH complex in basal transcription. The mutations that cause XP, in contrast, affect the role of TFIIH in NER. The mutations that underlie XP-CS appear to affect *both* basal transcription and NER, demonstrating that all of these diverse phenotypes are expressed as part of a continuum.

DNA Repair Defects and Mutagens Define Two Steps Towards Genetic Instability

Although rare, XP and the related NER-associated syndromes illuminate several central principles of cancer genetics. XP-associated cancers, like many types of cancer, can be caused by distinct mutations in different genes that function in a common biochemical pathway. XP gene mutations exhibit pleiotropy. As we have seen, some cancer gene mutations can cause phenotypes entirely unrelated to cancer.

It is revealing that XP carriers, individuals heterozygous for XP gene mutations, are asymptomatic and do not have a measurably increased risk of cancer. By definition, XP genes are tumor suppressor genes because they lose function as a result of mutation and cause cancer. But the tumor suppressor genes that underlie XP are clearly dissimilar from the classical tumor suppressor genes.

The cancer syndromes presented in Chapter 3 are all autosomal dominant. In the more prevalent of these, the inheritance of a single mutated tumor suppressor gene allele confers a significant risk of cancer. In the case of germline *RB* and *APC* mutations, this risk is close to 100%. Other tumor suppressor genes, such as the HNPCC genes, are incompletely penetrant, yet still confer cancer susceptibility. In contrast, the penetrance of a single mutated XP allele in the germline would seem to be close to nil.

Classical tumor suppressor alleles become unmasked and exert their phenotypic affects when expanding cell clones lose the remaining wild type allele, a process known as LOH (see Chapter 3). Apparently, LOH at the XP loci is alone insufficient to generate a level of genetic instability that would elevate the overall risk of cancer (see Fig. 4.12). Genetic instability in NER-deficient cells is dependent not only on LOH at the XP loci, but also on the mutational effects of UV light.

To understand the rate-limiting role of the environment in XP, it is perhaps useful to compare XP and HNPCC. Although cancers in both diseases arise as a result

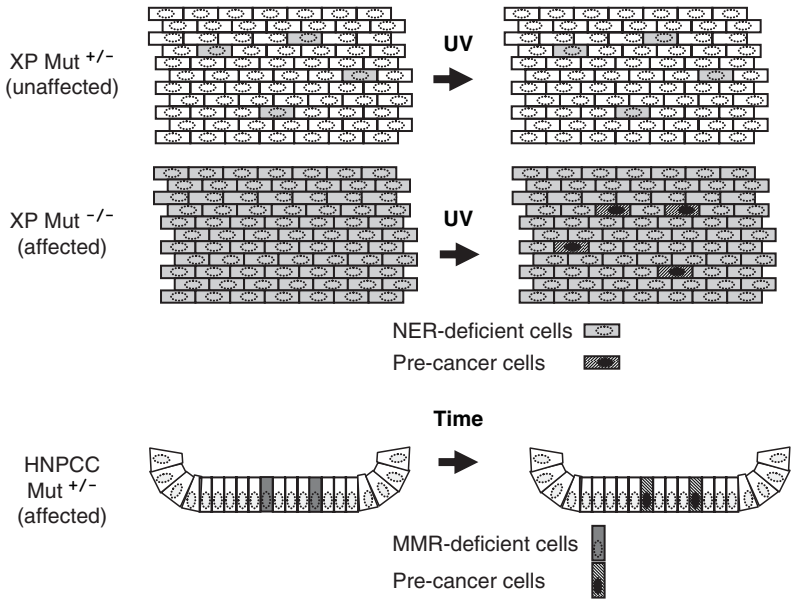

Fig. 4.12 DNA repair defects and environmental mutagens. Patients affected with XP are homozygous for XP gene mutations. Every cell in these patients is NER-deficient (shown in gray). Following UV exposure, mutations cause a small but significant minority of proliferating cells in the skin to become cancer precursors (hatched cells). In heterozygous XP carriers, wild type XP genes are presumably lost, via LOH, at a low rate. Because (1) the total number of NER-deficient cells is very small, and (2) only a small proportion of NER-deficient cells become cancer precursors after UV exposure, XP carriers are not at an elevated risk of cancer. The situation differs in HNPCC, a disorder with an autosomal dominant mode of inheritance. In individuals heterozygous for HNPCC mutations, LOH in the cells of the colon crypt causes MMR-deficiency (gray cells). MMR-defective cells acquire mutations as a byproduct of normal cell proliferation and have significant potential to develop into cancers in later generations

of defective DNA repair, XP contrasts starkly with HNPCC, an autosomal dominant disorder that causes high rates of cancer in heterozygotes. What is the basis for this difference? NER processes DNA lesions caused by the environment, while MMR processes misincorporated bases that arise during normal DNA replication.

In the case of HNPCC, genetic instability occurs immediately and in every proliferating cell that sustains LOH at the affected locus. In XP heterozygotes, loss of the wild type allele would not invariably lead to a higher mutation rate. Rather, genetic instability is only manifest in the presence of an environmental mutagen. That the increased occurrence of mutations in NER-deficient cells requires an exogenous component constitutes an extra step towards the acquisition of genetic instability. It appears that this extra step is rate-limiting in XP heterozygotes, which is likely to be the reason that the incidence of cancer in this population is close to normal.

Defects in DNA Crosslink Repair Cause Fanconi Anemia

Fanconi anemia (FA) is a recessive syndrome that features a predisposition to bone marrow failure and cancer. Hematological disease appears at a median age of 7 years. FA patients develop both liquid and solid tumors at greatly elevated rates. By age 40, the risk of developing a neoplasm is approximately 30%. The most prevalent type of cancer associated with FA is acute myelogenous leukemia (AML), which develops at a median age of 14 years. FA patients are also prone to solid tumors, including head and neck and gastrointestinal carcinomas. Solid tumors occur in FA patients at a median age of 26 years. The predisposition of FA patients to solid tumors is particularly striking because of the young age at which they occur, as compared to the general population. FA patients have a markedly reduced life expectancy, with median survival estimated at 23 years.

Cultured cells from FA patients exhibit elevated levels of spontaneous chromosome breaks and deletions. FA cells are also highly sensitive to agents that cause DNA crosslinks, such as the carcinogen diepoxybutane (DEB) and the chemotherapeutic drug mitomycin C (MMC). It is believed that the failure to normally repair and resolve DNA crosslinks directly leads to double-strand DNA breaks. Agents that cause chromosome breaks are known as *clastogens*. The spontaneous chromosome breakage and sensitivity of FA cells to clastogens represents a unique form of genetic instability. It is believed that this form of genetic instability is a direct cause of the cancer seen in these patients, but the specific mechanism of mutagenesis remains incompletely understood.

FA is a highly heterogeneous disorder. FA patients present with diverse clinical features that include congenital abnormalities that can affect any major organ system. First recognized and described by the pediatrician Guido Fanconi in 1927, FA is most commonly characterized by an abnormal reduction in the number of red blood cells, white blood cells and platelets in the blood, a condition known as *pancytopenia*. Other common features include short stature, hyperpigmentation, skeletal malformations, and urogenital abnormalities. Vertebral anomalies, the absence or closure of the anus, fistulae in the esophagus or trachea, limb and skeletal malformations, renal, gastrointestinal and cardiac abnormalities have all been associated with FA. Some patients do not exhibit any congenital abnormalities, or exhibit minor malformations that can be easily overlooked by pediatricians. In such patients, the diagnosis of FA is made after the appearance of hematologic disease. Because the presentation of FA is highly variable, a correct diagnosis may be difficult to make on the basis of clinical manifestations alone. In many cases, FA patients have been identified by virtue of their relatedness to previously diagnosed FA patients. The genetic analysis of FA has been significantly hampered by these inherent difficulties in ascertainment. The assessment of chromosome breakage after treatment with DEB or MMC can serve as a unique and highly useful laboratory test to definitively diagnose FA.

The clinical heterogeneity of FA is partly the result of underlying genetic heterogeneity. Like XP, FA can be subcategorized into complementation groups.

The FA complementation groups are defined by their ability to cross-complement the cellular hypersensitivity to the clastogenic effects of DNA crosslinking agents. Twelve complementation groups (designated FA-A, -B, -C, -D1, -D2, -E, -F, –G, -I, -J, -L and -M) have been thus identified. The majority of the genes mutated in these groups have been cloned.

The FA genes were identified by several related strategies that are similar to those used to clone the XP genes. Cells from the FA-C group, for example, were able to be complemented with pooled clones from a human cDNA library. From these pools, individual clones that could complement the chromosome breakage phenotypes of FA-C cells were identified. The gene corresponding to these complementing cDNAs was designated *FANCC*, and was the first FA gene identified. Analysis of *FANCC* in FA-C patients revealed frameshift, splicing, missense and truncation mutations. Similar approaches, combined with positional cloning, have been used to clone and validate *FANCA, FANCB, FANCE, FANCF* and *FANCG* as the genes corresponding to the FA-A, -B, -E, -F, and –G groups, respectively. The group initially designated FA-D has since been determined to be heterogeneous, with mutations in the genes *FANCD1* and *FANCD2* occurring in subsets of the FA-D group. FA-A represents the largest group of patients, accounting for approximately 65% of all FA cases.

The molecular cloning of the gene for the FA-D1 complementation group led to the discovery that *FANCD1* is identical to *BRCA2*. Therefore, *BRCA2* mutations can cause two distinct diseases. Germline inheritance of monoallelic *BRCA2* mutations causes familial breast and ovarian cancer susceptibility (an autosomal dominant syndrome), while biallelic germline mutations in *BRCA2* cause FA (an autosomal recessive syndrome).

FA is significantly more common than the NER syndromes. Because carriers are unaffected (with the exception of carriers of mutated *BRCA2*), and because of the difficulties in case ascertainment, the frequency of FA alleles in the general population has been difficult to accurately determine. It has been estimated that as many as 0.5% of the general population may be heterozygous for an FA gene mutation. The FA allele frequency is about 1% among individuals of South African Afrikaans or Ashkenazi Jewish descent, as a result of founder effects. For example, an A → T splice site mutation in *FANCC*, designated c.711 + 4A > T, is unique to FA patients of Ashkenazi ancestry. The incidence of FA syndrome is particularly high in ethnic groups in which consanguineous marriages are traditionally common.

The overall function of the FA gene-encoded proteins is not known in great detail. Biochemical analysis of these proteins has revealed multiple interactions with proteins that are known to be involved in DNA repair. Indeed, BRCA2 had previously been implicated in the repair of DNA double-strand breaks and DNA interstrand crosslinks. The association of FA proteins with sites of DNA damage and repair and the cellular sensitivity of FA cells to interstrand crosslinks provide strong evidence that FA results from DNA repair defects. The role of FA proteins in DNA repair and their activation by DNA damage will be described in Chapter 5.

FA generally exhibits an autosomal recessive mode of transmission, and accordingly, most FA genes are located on the autosomes. However, the gene mutated in

the FA-B group, *FANCB*, was found to be located on the X chromosome. This means that FA-B must exhibit a unique mode of transmission that, interestingly, had not been detected prior to the identification of the underlying genetic defect. Male carriers harbor hemizygous *FANCB* mutations and are therefore invariably affected with the disease. In females, the X-chromosome is randomly inactivated early in development. Normally, females are composed of cells that represent a *mosaic* with respect to the X-chromosome that is inactivated. Female carriers of X-linked *FANCB* mutations would therefore be expected to express the mutant protein in one half of their cells and thus partially express a mutant phenotype. In fact, these individuals appear to be perfectly normal. Cellular and molecular analysis has revealed that *FANCB*-mutant female carriers have a markedly reduced level of mosaicism as compared to females that do not carry a *FANCB* mutation. How might this occur? Cells that express mutant FANCB appear to be more prone to spontaneously undergo apoptosis, and are therefore at a significant proliferative disadvantage as compared with wild type cells. It would appear that, after X-inactivation occurs, cells that express mutant FANCB are outcompeted by the cells that express wild type FANCB (see Fig. 4.13). In the female mutant *FANCB* carrier, the majority of somatic cells are apparently derived from the embryonic precursors in which the

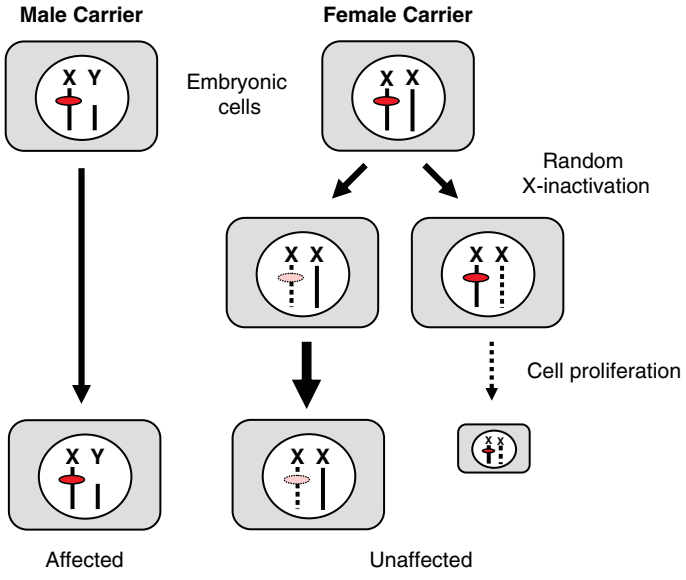

Fig. 4.13 X-linked inheritance of *FANCB* mutations. Male carriers of mutant *FANCB* (shown in red) are hemizygous for this recessive allele, and are therefore affected by Fanconi Anemia. In females, each X chromosome is subject to random inactivation. However, the cells derived from precursors that had inactivated the wild type *FANCB* allele are apparently at a significant proliferative disadvantage during subsequent phases of development. The large majority of cells in the developed female carrier are derived from the embryonic precursors in which the mutant *FANCB* allele had been inactivated. The female *FANCB* mutant carrier is therefore unaffected

copy of the X-chromosome harboring the *FANCB* mutation has been inactivated. This unusual mode of transmission has important implications for the genetic counseling of FA families in which males are exclusively affected.

In contrast to female carriers of *FANCB* mutations, who lose mosaicism during development, about one quarter of FA patients with autosomal mutations have been found to gain mosaicism. In such patients, two distinct populations of blood cells can be detected: one population with a marked chromosome-break phenotype induced by DEB or MMC treatment, and one that is phenotypically normal. In such mosaic patients, a proportion of blood cells have apparently reverted to a normal phenotype. There are two known mechanisms for reversion. In some cases, reversion is a result of mitotic recombination (see Fig. 4.14). Many FA patients are *compound heterozygotes*, in whom two distinct mutations are present in the maternal and paternal FA alleles. Intragenic recombination that occurs during mitosis can result in the transmission of both mutations on the same allele. Another mechanism of reversion is *gene conversion*, which occurs upon introduction of compensatory mutations (see Fig. 4.15). It has been demonstrated that frameshift mutations in FA genes can be compensated for by somatically acquired mutations that restore the correct reading frame and thereby revert to wild type function. It is unclear whether reversion of FA mutations in subpopulations of blood cells is sufficient to alter the course of the disease or significantly affect the prognosis.

While the genetic heterogeneity among the FA complementation groups contributes to the overall clinical heterogeneity of the disease, other factors additionally affect the expression of the varying disease phenotypes. Significant phenotypic variation has been reported within families. Even monozygotic twins have been found to be discordant in their expression of congenital abnormalities. These

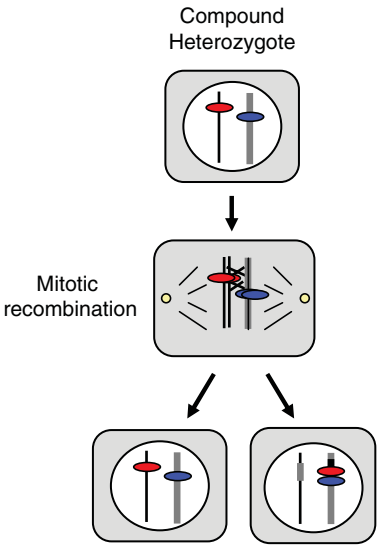

Fig. 4.14 Reversion of FA phenotypes by intragenic recombination. An FA-affected individual with compound heterozygous FA mutations (shown in red and blue) has two inactivated FA alleles. During mitosis, crossing over leads to an exchange between chromatids on the maternal and paternal chromosomes. A recombination event puts both FA mutations on the same chromosome, while the other chromosome reverts to wild type. The daughter cell that inherits the revertant allele is phenotypically normal. The patient exhibits mosaicism

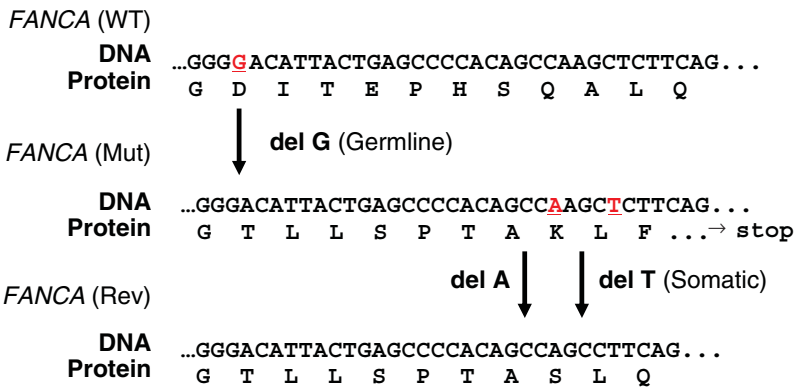

Fig. 4.15 Reversion of FA mutations by gene conversion. The germline deletion of a single G residue at position 1615 (underlined in red) in an FA patient disrupts the open reading frame of *FANCA (FANCA* (Mut)) and leads to the premature truncation of the encoded protein. Two somatic deletions occurring downstream of the germline mutation cause a shift back to the wild type reading frame *(FANCA* (Rev)) and restore function of the encoded protein. (Waisfisz et al. *Nature Genetics* **22,** 379–383 (1999).)

studies conclusively demonstrate that while FA gene mutations are the cause of FA, the specific features of the disease can be shaped by unique genetic and environmental factors.

A Defect in DNA Double-Strand Break Responses Causes Ataxia-telangiectasia

Ataxia-telangiectasia (AT) is an autosomal recessive syndrome characterized by hypersensitivity to ionizing radiation and a predisposition to cancer, most commonly in lymphoid tissues. Major clinical features of this disorder include: (1) a progressively disabling loss of muscle coordination that underlies a gait abnormality known as cerebellar ataxia, (2) an inability to follow an object across the visual fields, a symptom known as oculomotor apraxia, (3) dilated groups of capillaries, known as telangiectasia, which cause elevated dark red blotches on the skin and eyes, and (4) humoral and cellular immunodeficiencies that predispose affected patients to frequent infections. AT patients typically exhibit high levels of a serum protein known as α-fetoprotein, believed to be a suppressor of immune function. Unlike XP and FA, the AT syndrome is phenotypically homogeneous, and thus varies little from family to family.

A clue into the cellular basis of AT was provided by the observation that AT patients are highly sensitive to the effects of ionizing radiation, which is often employed as cancer therapy. Therapeutic doses of radiation are well tolerated by

most cancer patients, but cause serious and often life-threatening complications in AT patients. Ionizing radiation and drugs that mimic the effects of radiation impart several types of cellular damage, predominant among these are double-strand DNA breaks. Double-strand DNA breaks present a significant challenge to proliferating cells and are highly lethal when unrepaired.

Several aberrant responses to radiation can be observed in cultured, AT patient-derived cells. AT-associated radiosensitivity can be observed by measuring the proportional survival of irradiated cells (see Fig. 4.16). AT cells also have a characteristic defect in the regulation of DNA replication after irradiation. Normally, cells transiently pause DNA replication in progress at the time of radiation exposure. AT cells are defective in this radiation response and fail to pause replicative DNA synthesis– a phenomenon known as *radioresistant DNA synthesis* (RDS). RDS can be directly quantified by measuring the uptake of [^{3}H]- thymidine within a timed interval following radiation exposure. It is not clear whether RDS and reduced survival are causally related, but these two in vitro responses nonetheless facilitate the quantification of cellular AT phenotypes.

The initial experimental strategies to determine the genetic cause of AT were similar to those employed in the search for the NER genes. It was reported that the

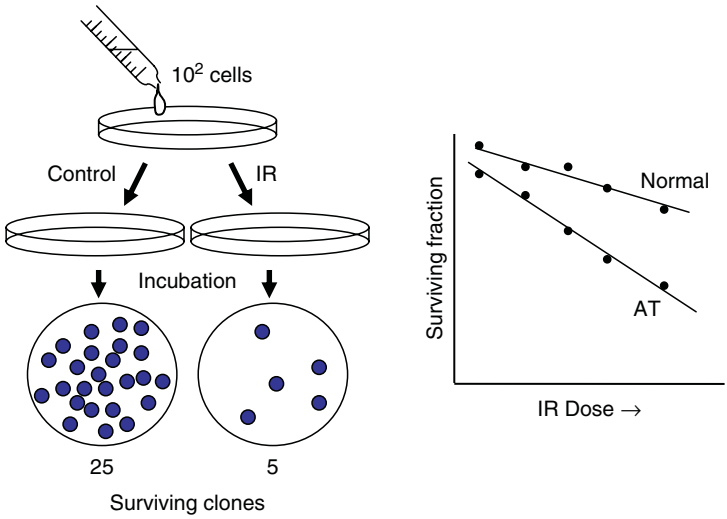

Fig. 4.16 Cells derived from AT patients exhibit reduced clonogenic survival following exposure to ionizing radiation. Cells – typically fibroblasts – are seeded to culture dishes and exposed to measured doses of ionizing radiation (IR). In this example, 100 cells are plated to multiple plates (left panel). Plates are incubated following treatment, allowing the surviving cells to proliferate and form colonies. Only a fraction of the original cell population is clonogenic, and thus forms colonies. On the untreated (control) plate, 25 clones (shown in blue) are visible after seeding and incubation; the *plating efficiency* of the original cell population is 25%. The plate treated with IR contains a reduced number of clones, reflecting reduced clonogenic survival. At this single dose of IR, the *surviving fraction* is 0.2. AT-derived cells are more sensitive to IR than are cells from normal individuals, across a wide dose range (right panel)

RDS phenotype of AT could be complemented by cell fusions. These data were used to categorize patients into a number of distinct complementation groups, similar to the manner in which XP patients were categorized by the complementation of UDS. However, subsequent efforts to clone AT-associated genes by complementation were unsuccessful. The relevance of the originally defined AT complementation groups remains unclear. Similarly, attempts were made to complement the radiosensitivity phenotype of AT cells by gene transfer. While numerous genes were isolated by this approach, none of these was found to be defective in AT patients. In retrospect, it would appear that the rescue of the RDS and radiosensitivity phenotypes in cultured AT cells did not reflect underlying genetic defect that causes AT but rather were artifacts of the methods employed.

Ultimately, the genetic basis of AT was revealed by linkage to a region on chromosome 11 by Richard Gatti and coworkers. This discovery guided subsequent positional cloning efforts. In 1995, a collaborative group led by Yosef Shiloh identified a single gene on chromosome 11q22–23 that was mutated in the germline of AT patients. This gene was designated *ataxia telangiectasia mutated*, or *ATM*.

ATM is a large gene composed of 65 coding exons. The mutations in *ATM* that cause AT are diverse and distributed throughout the *ATM* coding region. There are only a few relative mutational hotspots; many mutations are unique. The most common types of *ATM* mutations are single base substitutions and short deletions. Of the single base substitutions in *ATM*, more than one third are nonsense mutations. Therefore, the majority of *ATM* mutations result in the truncation of the open reading frame. AT patients are most commonly *compound heterozygotes*, and thus harbor two different *ATM* mutations.

AT patients have an approximately 100-fold increased lifetime risk of developing cancer. Approximately 85% of cancers that arise in AT patients are leukemias and lymphomas. The care of AT patients has improved in recent years and resulted in an improvement of the average life span. In the older cohort of AT patients, significant numbers of solid tumors such as breast cancers and melanomas have begun to be observed.

The incidence of AT has been estimated at 1 in 40,000 live births. About 1% of the general population carries a mutant *ATM* allele. Cells from heterozygous carriers have been demonstrated to partially exhibit some of the cellular defects of AT. Despite these subtle cellular abnormalities, *ATM* mutation carriers appear to be otherwise phenotypically normal. Shortly following the cloning of *ATM* and the discovery of *ATM* mutations, several studies suggested that carriers have an increased risk of cancer, particularly breast cancer. More recently, large epidemiological studies have provided conclusive evidence that single, heterozygous ATM mutations cause breast cancer susceptibility. Conferring a relative risk of approximately two, mutant ATM alleles are low-penetrance breast cancer genes.

AT is relatively homogenous in its clinical presentation, as is often the case with mongenic disorders. Nonetheless, affected individuals characteristically exhibit many diverse disease phenotypes. The spectrum of disease features suggests that loss of ATM function affects tissues in different ways. For example, the Purkinje cells of the cerebellum degenerate and migrate abnormally in the absence of ATM

activity, while the thymus fails to develop beyond the embryonic stage. That the neurological, immunological and neoplastic characteristics of AT are all attributable to a single genetic defect demonstrates an extraordinary degree of pleiotropy. The pleiotropy of *ATM* mutations suggests a broad role for DNA break responses in normal human physiology.

ATM encodes a large protein kinase that is rapidly activated at the site of DNA breaks. Activated ATM associates with a multiprotein complex known as the MRN complex, and with the protein encoded by *BRCA1*. The MRN complex binds DNA and has multiple biochemical properties that include the cutting, unwinding and bridging of the ends of the damaged double helix. BRCA1 protein is involved in non-homologous DNA end joining (see Chapter 1). It appears that the MRN complex and the BRCA1 protein are required for the efficient recruitment and possibly the retention of ATM at double-strand DNA break sites. The molecular mechanism of DNA break-dependent activation of ATM will be described in Chapter 5.

The MRN complex contains three proteins that are encoded by *MRE11*, *RAD50* and *NBS1*. Germline mutations of each of these genes have been found in individuals with rare autosomal recessive syndromes that clinically overlap with AT. *MRE11* mutations cause the Ataxia Telangiectasia-like disorder (ATLD), mutations in *NBS1* cause Nijmegen breakage syndrome (NBS), and mutations in *RAD50* have been found in a single individual with an NBS phenotype. While ATLD has not been associated with an increase in cancer risk, individuals with NBS are prone to leukemias, melanomas, and cancers of the prostate, breast and ovary.

The importance of the DNA damage response genes to the suppression of cancer is underscored by the numerous tumor suppressors that are functionally linked to ATM and the DNA damage response. The proteins of the MRN complex are important for ATM activation and the efficient transduction of DNA damage signals to downstream target molecules. Among these are the FA proteins and p53, which is encoded by the most prevalently inactivated tumor suppressor gene.

Bloom Syndrome Features Hyper-recombination

Among all of the cancer syndromes described in this chapter, none more emphatically highlights the causal relationship between genetic instability and cancer than does Bloom Syndrome (BS). Affected individuals exhibit both a readily observable defect in the maintenance of chromosomes and a pronounced predisposition to develop common forms of cancer.

BS patients develop cancer at a higher rate than any other genetically defined group of individuals. The types of cancer and the sites at which they occur are similar to those in the general population. About 30% of BS patients develop leukemias and lymphomas, and a similar proportion develops carcinomas of various types. Typically, the liquid tumors arise in younger patients, whereas the carcinomas develop later in life. Of the BS patients who develop malignant disease, about 10% have more than one primary cancer. BS patients develop cancers at an early age,

with a mean age of onset of 25 years. Other phenotypes associated with BS, include small body size, characteristic facies and voices, sun-sensitivity, immunodeficiency and, in males, infertility. Cancer accounts for a markedly reduced life expectancy for BS patients, the majority of whom die before age 30.

Proliferating cells isolated from BS patients exhibit several striking chromosomal abnormalities. Chromosomes from dividing blood lymphocytes exhibit numerous chromatid breaks, gaps, and structural rearrangements. These structural features occur spontaneously, in the absence of any environmental stimulus or clastogen. BS-associated cytological defects are quantitative in nature; similar features can also be seen in cells from normal individuals, but at a much lower frequency. The process that underlies this elevated level of gross chromosomal aberrations is an abnormally elevated rate of recombination between homologous chromosome regions. Recombination can occur between two chromosomes of a homologous pair, or intrachromosomally, between the sister chromatids of a single chromosome. These inter- and intra-chromosomal exchanges probably occur during S-phase, when chromosomes are replicated.

Intrachromosomal exchanges can be visualized by differentially labeling sister chromatids during S-phase and examining stained chromosomes during a subsequent metaphase (see Fig. 4.17). This type of analysis allows the extent of recombination to be quantified. BS cells exhibit a high level of sister chromatid exchanges (SCE). While cells from normal individuals typically exhibit fewer than 10 SCE/metaphase spread, BS-derived cells often exhibit from 60 to 90

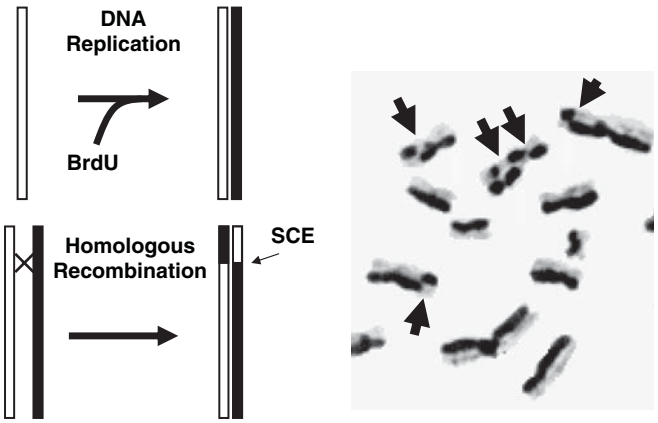

Fig. 4.17 Genetic instability in Bloom Syndrome. Sister chromatids can be differentially stained by incubating mitotically active lymphocytes for one cell cycle in the presence of the nucleotide analog bromodeoxyuridine (BrdU). The newly synthesized chromatid (shown in black), contains incorporated BrdU and is then photobleached and stained. Sister chromatid exchanges (SCE) resulting from homologous recombination can be visualized as alternating regions of light and dark staining. Bloom syndrome metaphase chromosomes (right panel) exhibit numerous SCEs (indicated by arrows)

SCE/metaphase. This elevated SCE frequency, known as the *high-SCE* phenotype, is highly diagnostic for BS. In addition, BS cells exhibit an elevated frequency of mutations that occur at the submicroscopic level, including point mutations and mutations at repeat sequences. By several criteria, the genome in BS cells is highly unstable.

Interestingly, the cells from some BS patients are found to vary in their phenotypic presentation. While the majority of lymphocytes from a given BS patient exhibit the high-SCE phenotype, a significant proportion might be found to be low SCE, and thus functionally wild type. These individuals were thus mosaic in respect to expression of the cellular BS phenotype. A similar type of mosaicism is observed in some patients with Fanconi Anemia. The explanation for mosaicism in FA is that compound heterozygotes, individuals who have different mutations on each of the two alleles, can create a normal allele by the process of homologous recombination (see Fig. 4.14). As the primary defect in BS is an increased rate of homologous recombination, it appeared probable that such a mechanism would explain the reversion of the BS phenotype observed in mosaic patients. Definitive proof of this hypothesis awaited the identification of the BS gene.

BS is a monogenic disease caused by the mutational inactivation of a tumor suppressor gene. Like many other tumor suppressor genes, the BS gene was cloned by virtue of its chromosomal location. However, the methods employed exploited several unique aspects of BS inheritance as well as the BS-associated cellular phenotype. The overall approach was guided by several unique insights that, in a stepwise fashion, narrowed the search for the gene of interest. The first step in determining the position of the BS gene was the complementation of the high-SCE phenotype by whole chromosome transfer. In 1992, it was demonstrated that the transfer of a normal chromosome 15 could suppress high SCE in BS-derived cells.

Although extremely rare, BS is significantly more prevalent in the Ashkenazi Jewish population than in other ethnic groups. Among non-Ashkenazi families, the disease arises most frequently as a result of consanguinity. Such affected individuals exhibit many regions of homozygosity. The location of the BS gene on chromosome 15 was localized to 15q26 by using polymorphic markers to determine the extent of homozygosity, a technique that came to be termed ‘homozygosity mapping’. In Ashkenazi Jewish families, tight linkage was demonstrated between the BS allele, a gene designated *FES* located at 15q26.1 and several microsatellite repeat sequences.

The precise location of the BS gene was deduced by detailed genetic analysis of the BS-derived cell lines that were low-SCE revertants. In these lines, crossover events had reduced the compound heterozygous BS locus to homozygosity, thereby eliminating the high-SCE phenotype. Loci distal to the BS gene in these cell lines were homozygous, while proximal loci retained the heterozygosity observed in patients’ constitutional DNA. This observation facilitated a third cloning strategy termed ‘somatic crossover-point mapping’, which involved the identification of the junction between homozygous and heterozygous regions. Because the yet-to-be discovered BS gene had apparently been restored to wild type by

recombination, it was expected that the homozygous/heterozygous junction must fall within the locus of interest. Using this strategy, a group led by James German localized the BS gene to a relatively short interval of just 250 kb, and cloned the gene in 1995. Thus, the hyper-recombinant phenotype of BS was successfully employed as a tool to pinpoint the disease locus. The BS gene, designated *BLM*, is composed of 22 exons and spans approximately 100 kb on chromosome 15.

The role of *BLM* in BS was confirmed by the presence of mutations that segregate with the disease phenotype. A variety of mutations affect the *BLM* open reading frame, including missense and nonsense mutations, small indels and splice site mutations. Most of the *BLM* mutations found in BS patients result in premature truncation of the open reading frame and predicted inactivation of the protein product. Sequence analysis of *BLM* in the Ashkenazi Jewish population definitively confirmed the existence of a founder effect. The carrier rate in this population is 1%. Also confirmed by DNA sequencing was the prediction that the individuals exhibiting revertant cell populations were compound heterozygotes.

The *BLM* gene encodes an enzyme that belongs to a previously identified family of highly conserved DNA and RNA helicases. These enzymes catalyze the ATP-dependent unwinding of duplex nucleic acids, a process that is essential for basic cellular processes including DNA replication and repair, RNA transcription and protein translation. The protein encoded by *BLM* most closely resembles a helicase subfamily known as RECQ, named after the prototypic *RECQ* gene in the bacterium *E. coli*. Analysis of RECQ homologs in model organisms has yielded significant clues as to what the specific functions of human RECQ helicases might be. Bacterial *RECQ* is required for recombination during conjugation and also for resistance to UV, which is a potent inhibitor of DNA replication. In the budding yeast *S. cerevisae*, mutants of the *RECQ* homolog *SGS1* feature slow growth, frequent chromosome missegregation and chromosome rearrangements, and defects in double-strand break repair. In the fruitfly *D. melanogaster*, RECQ mutants confer sensitivity to mutagens as well as a pattern of sterility that resembles that observed in human BS. The sterile phenotype in flies has been attributed to chromosome missegregation that occurs prior to meiosis. Experimental disruption of the *BLM* gene in mice and chicken cells results in a high-SCE phenotype that closely resembles the cytological defect in BS.

The cloning of *BLM* and the analysis of homologs in model organisms has revealed a critical role for RECQ helicases in the process of homologous recombination and in the maintenance of genetic stability. The extraordinary risk of cancer borne by BS patients demonstrates in dramatic fashion the role of genetic instability in the development of common types of cancer.

Aging and Cancer: Insights from the Progeroid Syndromes

Cancer is strongly associated with aging. While cancer strikes individuals at all stages of life, the overall incidence of the most common cancers clearly increases with age.

The relationship between cancer and aging is readily explained by the cancer gene theory. As described in Chapter 1, neoplastic cell clones iteratively expand and accumulate mutations. The development of neoplastic clones into cancer requires many generations of cell growth. In most tissues this process takes years or even decades. Consequently, cancers tend to disproportionately appear in older individuals. Inborn genetic instability alters the time frame in which tumors develop. Common among HNPCC, XP, FA, AT and BS is the incidence of cancer at a young age. By various mechanisms, the genetic mutations that underlie these diseases cause genetic instability that, in turn, accelerates the process of tumorigenesis.

As is the case with cancer, many of the cardinal signs of aging can also be related to the maintenance of a stable genome. Evidence to support a direct relationship between aging and genetic stability is derived from studies of a category of inherited diseases known as the *progeroid syndromes. Progeria*, or premature aging, can be caused by inborn genetic instability.

The most symptomatically striking progeroid syndrome, and the most intensively studied, is Werner syndrome (WS). WS patients prominently exhibit a prematurely aged appearance that develops during the second and third decades of life. Affected individuals develop normally, but as young adults they develop grey hair, hyperpigmentation and other age-associated skin changes, and a hoarse voice. Individuals with WS appear 20–30 years older than their chronologic age (see Fig. 4.18). Many

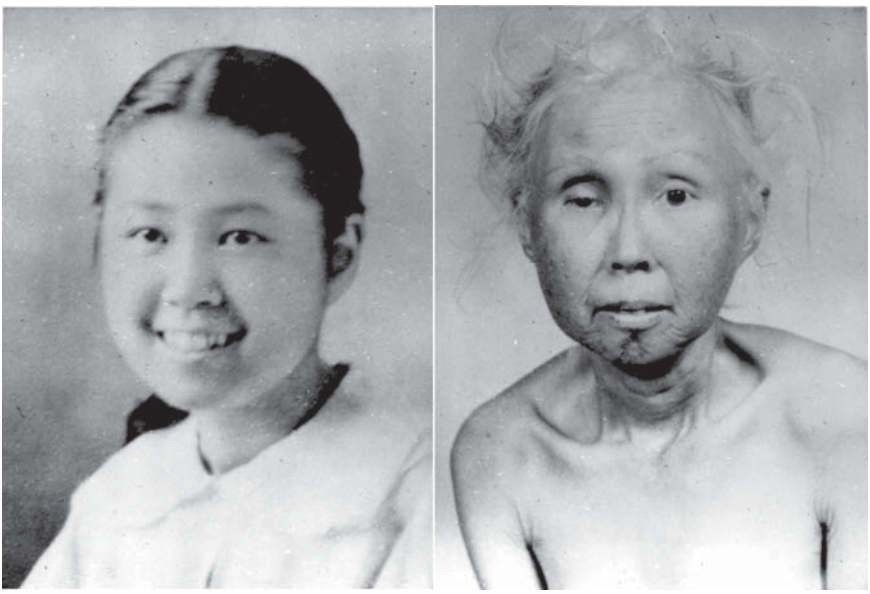

Fig. 4.18 Premature aging in Werner Syndrome. A Japanese woman with WS at age 15 (left) and at age 48 (right). (Epstein, C. J. et al. *Medicine* **45**, 177 (1966).)

disease states that are strongly associated with aging occur prematurely in WS patients. These typically include ateriosclerosis, cataract formation, osteoporosis and diabetes mellitus. WS patients are also at a highly elevated risk of cancer. Malignancies occurring in WS include a wide range of carcinomas and sarcomas. The average life span of WS patients is 47 years, and cancer is the cause of death in 80% of these individuals. WS is inherited in an autosomal recessive pattern; affected families often have a high rate of consanguinity.

While WS patients develop a broad range of cancers, they are not at increased risk for all common types of cancer. Rather, WS exhibits a selective increase in some relatively rare cancers. For example, while sarcomas occur infrequently in the general population, they represent approximately one half of the cancers that arise in WS. Conversely, several of the most common cancers in the general population, such as prostate cancer, do not occur at elevated rates in WS patients. WS thus does not exactly recapitulate all of the changes in cancer incidence that occur with normal aging. Similarly, WS does not exactly mimic the process of aging in its entirety. For example, the dermis of WS patients exhibits severe pathology that similar to that seen in normal aging, while the immune system appears to be unaffected. WS appears to mimic individual components, known as *segments*, of aging. To highlight this distinction, WS and the other diseases of premature aging are sometimes termed *segmental progeroid syndromes*.

Cells from WS patients show several characteristic phenotypes, including extensive chromosomal deletions, chromosome fusions, elevated rates of homologous recombination, a prolonged S-phase of the cell cycle, and defects in the DNA/protein structures at the ends of chromosomes, known as *telomeres*. That WS cells feature defects in telomere maintenance was a particularly provocative finding, as the shortening of telomeres had previously been implicated in the aging process. Telomeres are composed of unique repetitive DNA elements that are found nowhere else in the genome. Young individuals tend to have longer stretches of these telomeric repeats. Older individuals, in contrast, tend to have shorter telomeres, as do some cultured human cells that cease to divide. The cellular phenotypes of WS suggest defects in both DNA repair and in telomere metabolism.

WS cells are hypersensitive to DNA damaging agents that cause DNA double-strand breaks, including ionizing radiation. The degree of radiosensitivity is significantly lower than that seen in cells from ataxia telangiestasia or Nijmegen breakage syndrome patients, suggesting a distinct molecular defect in WS patients.

The WS gene was cloned by a positional approach. Localization of the gene was initially guided by homozygosity mapping, similar to the approach employed in the identification of the Bloom syndrome gene. Analysis of linkage to a panel of polymorphic markers narrowed down the location to a 1.2 mb region on chromosome 8. In a final comprehensive effort, ten genes that lie in this interval were screened for mutations. A previously uncharacterized gene, subsequently designated *WRN*, was found to be mutated in affected individuals and obligate carriers, by a group led by Gerard Schellenberg. *WRN* spans a 140 kb region at 8p11–12 and is composed of

35 exons. The first mutations identified in *WRN* were single base substitutions that created nonsense codons or splice site defects. Thus, the mutations that cause WS lead to the premature truncation of the *WRN*-encoded protein.

Initial characterization of the *WRN* coding region revealed a striking homology to RNA and DNA helicases. Like *BLM*, WRN encodes a RECQ-related helicase that unwinds nucleic acids in the 3′→5′ direction. Unique among the RECQ family members, WRN exhibits an N-terminal 3′→5′exonuclease activity that catalyzes the degradation of DNA ends. Biochemical studies have shown that the helicase and exonuclease associated with WRN have specificity to similar types of DNA structures, suggesting that these two domains function coordinately. Such a helicase could conceivably function in a number of cellular processes. The cellular phenotypes of WS suggest that the WRN helicase is likely to participate in some aspect of intrastrand DNA-crosslink repair. The precise cellular role of the WRN helicase and the means by which its loss of function causes WS remains to be completely elucidated.

The clinical features of WS would suggest that cancer and aging are part of a common overall process, all inextricably linked to genetic instability. However, a broader examination of the progeroid syndromes shows that the predisposition to cancer and the other various characteristics, or segments, are clearly separable (see Table 4.5). Two of the rare nucleotide excision repair syndromes, Cockayne syndrome and trichothiodystrophy prominently feature segmental progeria, but not cancer predisposition. In stark contrast, xeroderma pigmentosum features a 1,000-fold increase in the risk of skin cancer, as well as a significantly increased risk of

Table 4.5 Overlapping and distinct phenotypes related to inborn genetic instability

Disease	Gene(s)	Cellular defect: encoded enzyme activity	Cancer	Progeria
Xeroderma pigmentosum	*XPA–XPF*	NER: various DNA binding proteins, DNA helicases, endonucleases	+++	–
XFE progeroid syndrome	*XPF*	NER: Endonuclease	Unknown	+++
Trichothiodystrophy	*XPB XPD*	NER: DNA helicases	–	++
Cockayne syndrome	*CSA CSB*	Transcription-coupled repair: coupling factors	–	++
XP-CS complex	*XPB XPD*	NER: DNA helicases	+++	++
Ataxia telangiectasia	*ATM*	Double-strand DNA break recognition: Ser/Thr kinase	+	+
Bloom syndrome	*BLM*	DNA repair: RECQ helicase (exo-)	+++	++
Rothmund-Thomson syndrome	*RECQL4*	DNA repair: RECQ helicase (exo-)	++	++
Werner syndrome	*WRN*	DNA repair: RECQ helicase (exo+)	+++	+++

various internal cancers, but is not associated with progeria. Both Bloom syndrome and ataxia telangiectasia are characterized primarily by cancer predisposition, but also feature mild phenotypes related to segmental premature aging.

Genetic defects that affect nucleotide excision repair can variably cause cancer predisposition, segmental aging, both or neither. In the interesting case of the *XPD* gene, different point mutations cause xeroderma pigmentosum (cancer predisposition only), trichothiodystrophy (progeria only), or the XP-CS complex (cancer predisposition and progeria). Similarly, distinct mutations in *XPF* cause xeroderma pigmentosum or XFE progeroid syndrome, a phenotypically profound disease first described in 2006 and based on the discovery of a single affected individual. Rothmund-Thomson syndrome, a heritable disease that exhibits significant clinical overlap with WS, is caused by mutations in another RECQ family member, *RECQL4*. Rothmund-Thomson syndrome patients are at a greatly increased risk of developing osteosarcoma. Interestingly, different mutations in the *RECQL4* gene have been shown to cause two other rare, autosomal recessive disease syndromes, neither of which appear to feature an increased cancer risk nor segmental progeria. As clearly exemplified by *XPD* and *RECQL4*, different alleles of the same gene can lead to dramatically distinct diseases.

Overview: Genes and Genetic Instability

Rare autosomal recessive diseases such as Bloom syndrome and Werner syndrome do not have significant impact on the overall health of the human population. While certainly devastating to the affected individuals and their families, such diseases are rare and the frequencies of the respective cancer genes are low. Furthermore, it does not appear that the genes causing any of the recessive DNA repair-associated syndromes are somatically mutated in sporadic cancers. Nonetheless, the study of diseases caused by homozygous mutations of DNA repair genes diseases has led to key insights into the relationship between DNA maintenance, aging and neoplasia. Unraveling the molecular basis of these unusual cancer syndromes has provided a unique view into the types of cellular processes that, when either partially or totally disabled, can lead to cancer. The identification of specific genes has directed vigorous research activity into how the information content of the genome is stably maintained.

Not all of the mutations that cause genetic instability are rare. Hereditary nonpolyposis colorectal cancer accounts for a small but significant proportion of the estimated 57,000 deaths from colorectal cancer that occur in the USA every year, and contributes to the incidence of several other types of cancer as well (see Chapter 6). Both Fanconi anemia and ataxia telangiectasia are caused by cancer genes that are present in up to 1% of the general population. As the consequences of carrier status remains to be conclusively determined, it remains a possibility that germline mutations in the Fanconi anemia and ataxia telangiectasia genes could increase the risk of cancer in many heterozygous individuals. Heritable genetic

instability is not a mere laboratory curiosity, but instead represents a quantifiable threat to public health.

The genes that ensure the maintenance of a stable genome are, by definition, tumor suppressor genes and fall into the subcategory known as caretakers. As described in detail in Chapter 3, the mutation of caretaker genes does not directly increase the proliferation of cell clones. Instead, caretaker genes prevent aberrant cell proliferation indirectly, by preventing the accumulation of mutations in other tumor suppressor genes and proto-oncogenes that modulate growth control.

Intriguingly, the most common form of genetic instability seen in cancer, aneuploidy, remains the most mysterious. A central goal of ongoing research is to systematically enumerate the many genetic and epigenetic changes to the cancer cell genome. This effort promises to provide insight into the elusive cause – or causes – of aneuploidy in common cancers.

Further Reading

Ahmed, M. & Rahman, N. ATM and breast cancer susceptibility. *Oncogene* **25**, 5906–5911 (2006).

Andressoo, J. O., Hoeijmakers, J. H. & Mitchell, J. R. Nucleotide excision repair disorders and the balance between cancer and aging. *Cell. Cycle* **5**, 2886–2888 (2006).

Cahill, D. P., Kinzler, K. W., Vogelstein, B. & Lengauer, C. Genetic instability and darwinian selection in tumours. *Trends Cell Biol*. **9**, M57–M60 (1999).

Cleaver, J. E. Cancer in xeroderma pigmentosum and related disorders of DNA repair. *Nat. Rev. Cancer* **5**, 564–573 (2005).

Duesberg, P. Does aneuploidy or mutation start cancer? *Science* **307**, 41 (2005).

de Laat, W. L., Jaspers, N. G. & Hoeijmakers, J. H. Molecular mechanism of nucleotide excision repair. *Genes Dev*. **13**, 768–785 (1999).

Ellis, N. A. Mutation-causing mutations. *Nature* **381**, 110–111 (1996).

Ellis, N. A. & German, J. Molecular genetics of Bloom's syndrome. *Hum. Mol. Genet*. **5**, 1457–1463 (1996).

Fearon, E. R. Human cancer syndromes: Clues to the origin and nature of cancer. *Science* **278**, 1043–1050 (1997).

Gatti, R. A. The inherited basis of human radiosensitivity. *Acta Oncol*. **40**, 702–711 (2001).

Gurtan, A. M. & D'Andrea, A. D. Dedicated to the core: Understanding the Fanconi anemia complex. *DNA Repair* **5**, 1119–1125 (2006).

Joenje, H. & Patel, K. J. The emerging genetic and molecular basis of Fanconi anaemia. *Nat. Rev. Genet*. **2**, 446–457 (2001).

Kastan, M. B. & Lim, D. S. The many substrates and functions of ATM. *Nat. Rev. Mol. Cell Biol*. **1**, 179–86. (2000).

Kipling, D., Davis, T., Ostler, E. L. & Faragher, R. G. What can progeroid syndromes tell us about human aging? *Science* **305**, 1426–1431 (2004).

Lengauer, C., Kinzler, K. W. & Vogelstein, B. Genetic instabilities in human cancers. *Nature* **396**, 643–69. (1998).

Margolis, R. L. Tetraploidy and tumor development. *Cancer Cell* **8**, 353–354 (2005).

Michor, F., Iwasa, Y., Vogelstein, B., Lengauer, C. & Nowak, M. A. Can chromosomal instability initiate tumorigenesis? *Semin. Cancer Biol*. **15**, 43–49 (2005).

Modrich, P. Mismatch repair, genetic stability, and cancer. *Science* **266**, 1959–1960 (1994).

Mohaghegh, P. & Hickson, I. D. DNA helicase deficiencies associated with cancer predisposition and premature ageing disorders. *Hum. Mol. Genet*. **10**, 741–746 (2001).

Rajagopalan, H. & Lengauer, C. Aneuploidy and cancer. *Nature* **432**, 338–341 (2004).
Rajagopalan, H., Nowak, M. A., Vogelstein, B. & Lengauer, C. The significance of unstable chromosomes in colorectal cancer. *Nat. Rev. Cancer* **3**, 695–701 (2003).
Shiloh, Y. The ATM-mediated DNA-damage response: Taking shape. *Trends Biochem. Sci*. **31**, 402–410 (2006).
Tomlinson, I. & Bodmer, W. Selection, the mutation rate and cancer: Ensuring that the tail does not wag the dog. *Nat. Med*. **5**, 11–12 (1999).

Chapter 5
Cancer Gene Pathways

What are Cancer Gene Pathways?

The previous chapters have described what cancer genes *are* and how they are acquired. But what do cancer genes *do*? How do the inactivation of tumor suppressor genes and activation of proto-oncogenes alter cell clones so that they evolve into cancers? The answer to this question has been revealed by the functional analysis of the proteins encoded by cancer genes and their wild-type counterparts. Cancer genes populate cellular pathways that control cellular proliferation and cell death.

To understand the roles that cancer genes play in the evolution of cancer cell clones, it is essential to appreciate the extent to which normal cells interact with and are controlled by their microenvironment. In normal regenerative tissues, cells grow and divide in response to myriad cues. The outer membrane of most cells is in direct contact with the extracellular matrix, with extracellular fluid and with neighboring cells. Various molecules traverse these routes carrying information. Diverse signals instruct cells to grow, to stop growing, to differentiate and mature or in some cases, to die. Cells in the proliferative compartments of the human body are literally bathed in signals.

Signals arise from many extracellular sources and are transmitted by several types of molecules. Three examples of signaling molecules illustrate the diversity of cell signaling pathways: cytokines, nitric oxide and hormones. Local signals are produced by activated inflammatory cells in the form of small water-soluble proteins known as *cytokines*. Cytokines bind specific receptors on the cell surface and are potent stimulators of cell growth as well as triggers of cell death. The free-radical *nitric oxide* (NO) is a small signaling molecule that has the ability to cross numerous anatomical boundaries and affect virtually every type of cellular function. NO is highly unstable in nature, and its levels can therefore rapidly change in a dynamic environment. *Hormones* secreted by the endocrine system are potent signals that affect the function of distant cells. For example, insulin secreted by the islet cells of the pancreas stimulates the uptake of glucose by hepatocytes in the liver, and the growth of adipocytes that compose fatty tissues. Cytokines, NO, and hormones are but three examples of signaling molecules that facilitate communication between one cell type and another, neighboring and distant.

F. Bunz, *Principles of Cancer Genetics*.

© Springer 2008

Regulatory signals also arise from intracellular sources. The cell growth and division cycle is highly regulated and closely monitored. The sequential stages of the cell cycle, particularly S-phase and mitosis, are exquisitely sensitive to damaged chromosomes and to stalled DNA replication forks. Factors that can break chromosomes and impair DNA replication include environmental agents (e.g. ultraviolet light) and metabolic aberrations (e.g. the depletion of nucleotide precursors). Interruption of cell cycle progression triggers vigorous cellular responses that strongly affect proliferation and survival.

Mutations in cancer genes affect the responses of cells to changes in their internal and external environments. These genetic alterations cause changes in growth and invasion that are the essence of the cancer cell. Cancer cells have several acquired phenotypes that facilitate their clonal outgrowth and thereby lead to cancer morbidity. These phenotypic traits include:

Independence from external growth signals. Normal cells require stimulation from the external environment to grow and divide. Cancer cells have lost this requirement.

Insensitivity to antiproliferative signals and hostile environments. Normal cells respond to specific signaling molecules that instruct them to halt the progression of the cell cycle and/or to differentiate. Normal cells are also sensitive to nonspecific environmental factors that impede cell growth by affecting chromosome integrity or cellular metabolism. Cancer cells have lost their sensitivity to both specific and nonspecific inhibitors of proliferation.

Insensitivity to death signals. A unique type of antiproliferative signal is one that triggers programmed cell death, also known as apoptosis. Normal cells respond to a variety of apoptosis-inducing signals, which are important in the maintenance of tissue homeostasis. Cancer cells have lost their sensitivity to molecules and cellular states that trigger apoptosis.

Tissue invasion and metastasis. Normal cells within solid tissues remain stably positioned with respect to their neighbors. Cancerous tissues lack homeostasis in part because cancer cells actively invade neighboring tissue structures and relocate to distant sites.

Cellular Pathways are Defined by Protein–Protein Interactions

Biochemical reactions within proliferating cells regulate growth, division, maturation and survival. These regulatory reactions serve to monitor both intracellular and extracellular environments, and create signals that allow cells to adapt to changing physiologic states. Cellular signaling pathways are composed of a series of biochemical reactions that transmit information between molecules. The resultant flow of information within the cell allows it to respond to a dynamic environment.

The pathways that dictate cellular physiology are numerous and complex in their relationships to one another. Consider an electronic microprocessor that contains numerous integrated circuits. Microprocessors receive many inputs in the form of electrical currents. Some of these currents are amplified, others are transformed. The

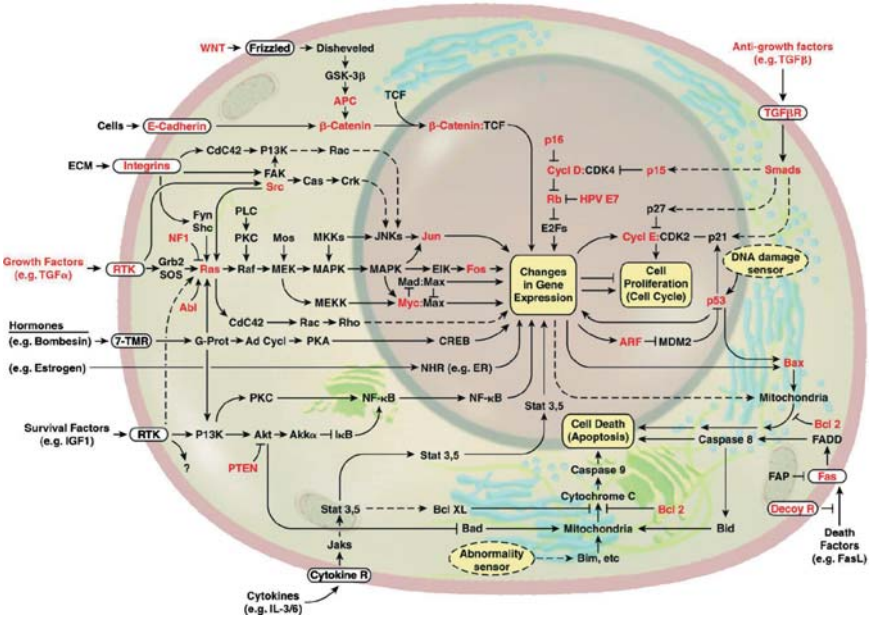

Fig. 5.1 Overview of cancer gene pathways. Human cells have highly integrated pathways that sense changes in the external and internal environments and trigger changes in growth. Numerous cancer genes populate these pathways. Cancer gene-encoded proteins shown in red. (Reprinted from Hanahan & Weinberg 2000. With permission from Elsevier.)

circuitry of the microprocessor integrates many inputs and generates an organized output. Like microprocessors, the signaling pathways of the cell are highly integrative and interconnected circuits that act as conduits of information (see Fig. 5.1).

Proteins are the primary nodes of the cellular pathways that are known to be defective in cancer. Proteins communicate with one another through direct, physical interactions. As a result of these types of interactions, one protein can be structurally and functionally altered by another.

The majority of proteins involved in cancer gene pathways are enzymes that catalyze the covalent modification of other proteins on specific amino acid residues. A covalent modification to an existing protein is known as a *posttranslational modification*. There are many ways in which proteins can be posttranslationally modified; several of most prevalent types of modifications are listed on Table 5.1.

The common covalent protein modifications are reversible. In all cases, the covalent attachment of the modification to a specific amino acid residue is catalyzed by one enzyme, and the reverse reaction is catalyzed by a second, distinct enzyme. Covalent posttranslational modifications can define two distinct states (see Fig. 5.2). Because the enzymes that add and remove modifications (depicted as ON and OFF enzymes respectively, in Fig. 5.2) are distinct, one state is usually strongly favored over the other, depending on which modifying

Table 5.1 Reversible protein modifications

Molecule	Size	Target	'ON' enzyme	'OFF' enzyme
Phosphate group–PO_3	79 Da	Ser, Thr, Tyr	Protein kinase	Protein phosphatase
Methyl group–CH_3	15 Da	Arg, Lys	Protein methyl-transferase	Protein demethylase
Acetyl group–$COCH_3$	43 Da	Lys	Protein acetylase	Protein deacetylase
Ubiquitin–polypeptide	8.5 kDa	Lys	Multiple sequential enzymes	De-ubiquitinase
Small ubiquitin-related modifier (SUMO) – polypeptide	10–11 kDa	Lys	Multiple sequential enzymes	SUMO isopeptidases

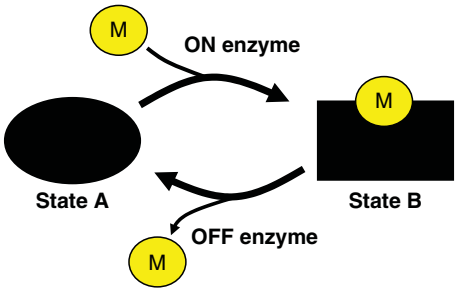

Fig. 5.2 The covalent modification of a protein reversibly alters its functional state. A hypothetical protein exists in two states, A and B. The transition from state A to B is mediated by the addition of a modifying group (M, shown in yellow) to an amino acid residue. This reaction is catalyzed by a hypothetical 'ON' enzyme. The reverse reaction, resulting in the removal of the modification and the transition from state B to state A, is catalyzed by a distinct 'OFF' enzyme. Either state A or state B can be the activated state, depending on the protein and the modifying molecule

enzyme is most active. The transition from one state to another can have several interrelated effects:

Structural effects. Most modifications cause a change in either the tertiary structure (the three-dimensional protein conformation) of a protein, or in the quaternary structure of a multiprotein complex.

Functional effects. The structure and function of any protein are inextricably interrelated. Modifications that change protein structure therefore usually alter protein function. In the case of proteins that are enzymes, modifications can change the activity of the catalytic domain, often by altering the ability of the substrate to bind. Modifications can either increase catalytic activity or decrease catalytic activity, depending on the protein, the location of the modification and the modifying group.

Localization. Modifications can affect the trafficking of a protein within the cell, and thereby affect its localization. Altered localization can affect access to interacting proteins, or to substrates.

Stabilization. Modifications can dramatically alter the half-life of a protein. Proteins can be either stabilized or destabilized as a result of modifications.

Individual Biochemical Reactions, Multistep Pathways, and Networks

The covalent modification of an individual protein is a single step in the series of biochemical reactions that define a *pathway*. Most pathways have several steps that involve multiple proteins and modifying molecules (see Fig. 5.3). When compared with a single biochemical reaction, a complex series of reactions has several added functional attributes. First, a higher order of organization significantly increases the extent to which a response to a stimulus can be controlled. Second, the multistep nature of most pathways allows a greater range of signal strength – known as the amplitude – than could be transmitted by a single reaction. Third, a multistep pathway allows a signal from a single location to have effects in multiple locations within the cell that may be physically distant. Some of the most extensively characterized pathways, for example, transduce signals from the cell surface to the nucleus.

Some pathways serve the purpose of amplifying a faint stimulus to produce a profound biochemical response. A notable example of this type of amplification is apoptosis. The interaction of two protein complexes on the cell surface is sufficient to generate

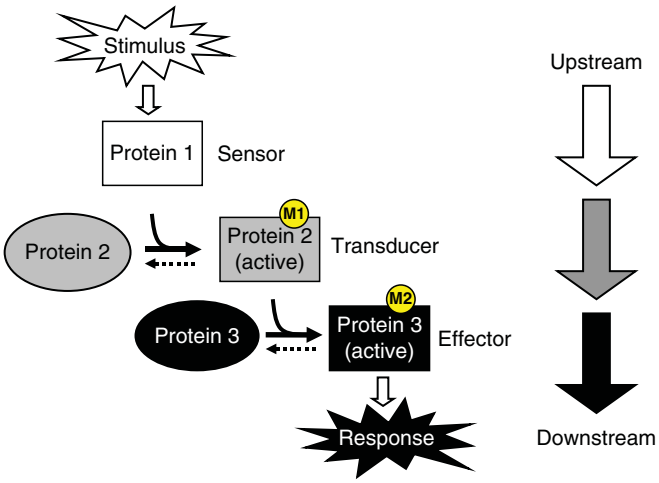

Fig. 5.3 A generic pathway. Several features that are common to many cellular pathways are shown. All pathways are characterized by directionality; signals are said to be transduced from an upstream stimulus to a downstream response. This hypothetical pathway is activated by a stimulus that activates a sensor (Protein 1). This sensor protein then adds a modification (M1, shown in yellow) to a second protein (Protein 2), causing it to become catalytically active. Once activated, Protein 2 serves as a transducer of the signal to a downstream effector (Protein 3) by catalyzing the addition of a modifying group (M2). Protein 3 then directly catalyzes a response. In many pathways, the intermediate transducer proteins can amplify upstream signals. The removal of a modifying group at any stage (dotted arrows) can result in the deactivation of the entire pathway. The hypothetical pathway shown has a single sensor, transducer and effector. Actual pathways can have multiple upstream and downstream proteins

a signal that is sequentially amplified and ultimately results in the proteolytic destruction of the cell. Pathways that sequentially amplify signals are known as *enzyme cascades*.

Multiple pathways can converge on a single protein (see Figs. 5.4 and 5.5). In such cases, the protein common to different pathways functions as a nodal point. Pathways can be either stimulatory or inhibitory. Because the nodes of a pathway integrate several types of proximal or *upstream* signals, they represent critical regulatory elements.

Other, more subtle interactions can also occur between pathways that, upon initial analysis, may appear to be unconnected. In such cases, two pathways will have distinct stimuli and distinct responses (see Fig. 5.6). Though such pathways are parallel in structure, the activation of one can sometimes, and under some conditions, positively or negatively affect the activation of the other. This type of regulation is often referred to as *crosstalk*.

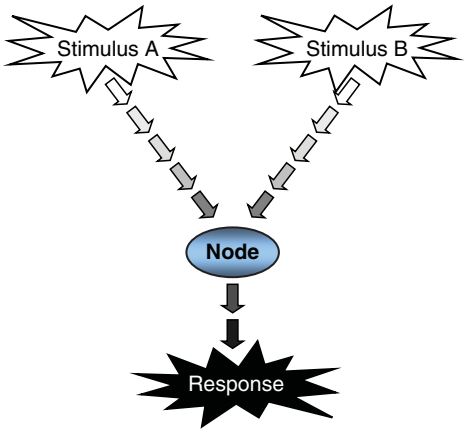

Fig. 5.4 Convergent pathways. Distinct upstream signals can lead to a common response. Here, two pathways, triggered by stimulus A and stimulus B, converge at a single point and join a common downstream pathway. Points at which pathways intersect are sometimes referred to as nodes

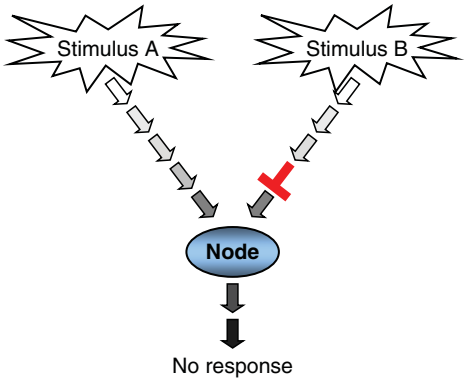

Fig. 5.5 Inhibitory and stimulatory pathways can converge. Some pathways can trigger reactions (shown in red) that inhibit downstream signaling events. In this example, the response that would be triggered by stimulus A is attenuated by the pathways activated by stimulus B

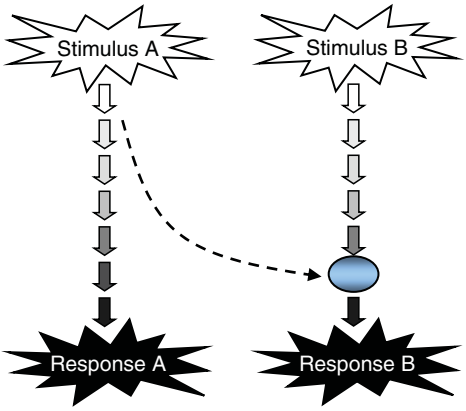

Fig. 5.6 Crosstalk between parallel pathways. In this example, Stimulus A leads primarily to Response A. Stimulus B leads to Response B via a distinct pathway. The 'A' pathway is interconnected (dotted line) with the 'B' pathway at a node. Stimulus A can thus affect Response B to some degree. Crosstalk can increase or decrease the signals transduced by pathways that are otherwise parallel in structure

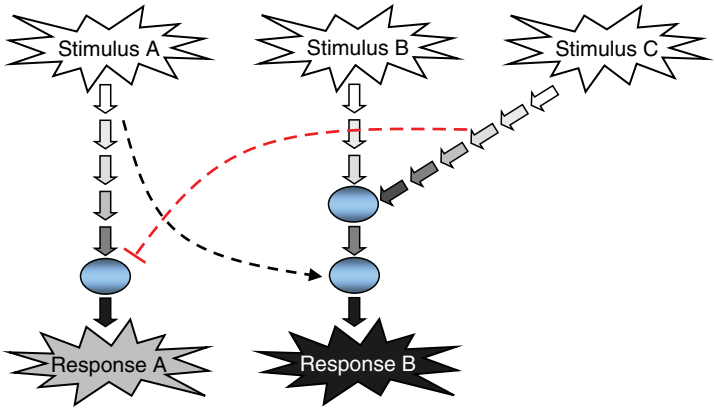

Fig. 5.7 Interconnected pathways form signaling networks. Multiple upstream signals affect multiple downstream responses. The activation of pathways that are influenced by crosstalk provides highly modulated signals that can effect nuanced responses. Shown is a simple multi-nodal network in which responses are stimulated by three activating pathways that are influenced by both stimulatory (black dashed line) and inhibitory (red dashed line) crosstalk. Responses A and B can thus be modulated with high precision

In effect, nodes link multiple individual pathways into functional signaling networks (see Fig. 5.7). The integration of multiple stimulatory and inhibitory pathways that are modulated by crosstalk creates a highly sensitive system with a large dynamic range. The extraordinary degree of connectivity that defines a network facilitates the finely tuned cellular responses to complex environmental stimuli.

As will be described in a later section, the p53 protein is a node of several critical pathways that regulate cell growth. It is certainly not coincidental that the *P53* gene is the most prevalently mutated gene in many types of cancer.

Protein Phosphorylation is a Common Regulatory Mechanism

The most widespread form of protein modification is phosphorylation. Phosphorylation and dephosphorylation of proteins are catalyzed by protein kinases and protein phosphatases, respectively. Kinases catalyze the transfer of the γ–phosphate group from adenosine triphosphate (ATP) to protein residues, while phosphatases catalyze the removal of this phosphate group. It is thought that up to 30% of the proteins encoded by the human genome variably contain covalently bound phosphate. The human genome encodes approximately 1,000 kinases and 500 phosphatases that mediate these transactions. The reversible phosphorylation of proteins affects virtually every cellular activity and function.

Until the 1970s, protein phosphorylation was viewed primarily as a specialized mechanism for the control of the pathways involving carbohydrate metabolism. The broader importance of protein phosphorylation only became apparent during the late 1970s and early 1980s, when roles of phosphorylation in diverse cell signaling pathways were discovered. Also appreciated at that time was that not all phosphorylated proteins are enzymes. Proteins that lend structure, organization and motility to cells are also targets of kinases and phosphatases. The reversibility and versatility of phosphorylation probably explains its evolutionary selection as a general mechanism of regulation.

A role for phosphorylation in cancer gene pathways was first discovered in 1978 by Ray Erikson, who demonstrated that the transforming protein encoded by the Rous sarcoma virus *V-SRC* gene encodes a protein kinase. Shortly thereafter, Tony Hunter demonstrated that the V-SRC protein catalyzes the addition of phosphate moieties to tyrosine residues, the first time such a catalytic activity had been observed. Abnormalities involving protein phosphorylation are now known to be among the most common defects found in cancer cells.

In addition to tyrosine, proteins can also be phosphorylated on serine and threonine residues (see Fig. 5.8). The addition of a phosphate group increases the molecular weight and the overall space occupied by an amino acid residue. Most importantly, phosphorylation changes the ionic charge of a residue so modified. Serine, threonine and tyrosine residues are neutrally charged; upon phosphorylation they become negatively charged.

The conformation of a protein is largely determined by interactions with neighboring water molecules. *Hydrophilic* regions of a protein tend to associate with water molecules, while *hydrophobic* regions tend to associate more closely with one another. The covalent addition or removal of negatively charged moieties lowers the hydrophobicity of the protein at that position, and thereby alters protein conformation.

The addition or removal of a phosphate moiety is a chemically simple modification that can have many functional consequences. Phosphorylation states can affect protein conformation and thereby increase or decrease biochemical activity. Phosphorylation and dephosphorylation can also control protein localization, protein stability and the direct interactions of proteins with other biomolecules.

Proteins that are regulated by phosphorylation/dephosphorylation are typically modified not at just one residue, but at several. Proteins can be phosphorylated at multiple sites by the same upstream kinase, or by kinases that belong to two

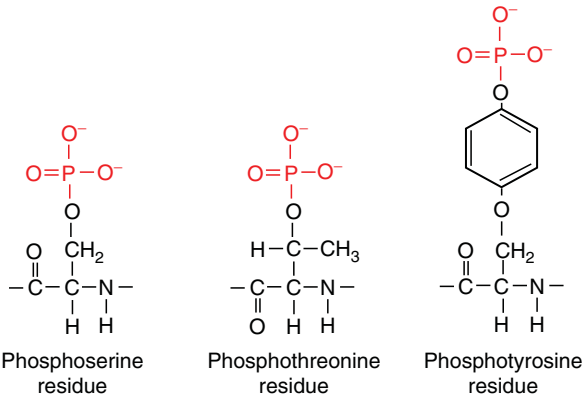

Fig. 5.8 Phosphorylated derivatives of serine, threonine and tyrosine. The addition of a phosphate group (red) adds a negatively charged moiety to a protein, altering its hydrophobicity and structure

different pathways. The phosphorylation of a single protein at multiple sites increases the extent to which is activity can be regulated by upstream signals. Additionally, multisite phosphoryation allows two different aspects of protein function, such as catalytic activity and half-life, to be separately regulated. The phosphorylation of a protein at one site can facilitate the phosphorylation or dephosphorylation of another site. In other words, phosphorylation states can exhibit *cooperativity*.

In some cancer gene pathways, protein phosphorylation/dephosphorylation functions essentially as a molecular on/off switch. Some of the key cell surface receptor proteins that control growth function in such a binary manner. In other pathways, phosphorylation functions as an extremely fine-tuned mechanism for regulating and coordinating the activity, timing and location of biochemical reactions. In the case of p53 activation, the phosphorylation of numerous residues and other posttranslational modifications allow p53 to integrate numerous types of upstream signals and serve to regulate many aspects of p53 activity.

Signals from the Cell Surface: Protein Tyrosine Kinases

Many signals that stimulate cell growth, cell division and cell death arise from the local microenvironment. Cells can sense the presence of both soluble and cell-associated signaling molecules known as *ligands*. Ligands, such as cytokines and hormones, represent extracellular signals that can be sensed and interpreted by receptors at the cell membrane. As a result of ligand–receptor interactions, signals from the cell surface are transmitted into a cell, and thereafter transduced throughout the cytoplasm and into the cell nucleus. Many cell surface receptors are protein tyrosine kinases. These signaling molecules process extracellular signals and, in response, activate intracellular cancer gene pathways. In many cases, protein tyrosine kinases are themselves encoded by cancer genes.

While protein phosphorylation is a very common posttranslational modification, only about 0.05% of all phosphorylated proteins are phosphorylated on tyrosine residues. Nonetheless, protein-tyrosine kinases are critical components of the signaling pathways that regulate cell proliferation. Many genes that encode protein tyrosine kinases are proto-oncogenes. Mutations that convert protein tyrosine kinase-encoding proto-oncogenes to oncogenes are commonly found in many types of cancer.

There are two broad categories of protein tyrosine kinases. *Receptor tyrosine kinases* (RTKs) are transmembrane proteins that span the cell membrane and channel signals from the outside of the cell to the cytoplasm. *Cytoplasmic protein tyrosine kinases* (CTKs) are intracellular enzymes that transduce signals throughout the cytoplasm and into the nucleus. Many CTKs are associated with the interior surface of the cell membrane, but do not span the lipid bilayer.

Of the roughly 100 protein tyrosine kinases that are encoded by the human genome, 58 are transmembrane proteins that function as receptors. On the basis of their structure, these RTKs can be grouped into 20 distinct families. The RTKs are highly specialized molecules that have evolved to mediate cell-to-cell communications and are particularly important during development. Accordingly RTKs are found exclusively in metazoans.

RTKs have a typical structure that defines their function (see Fig. 5.9). An extracellular domain is involved in ligand binding. A hydrophobic transmembrane domain

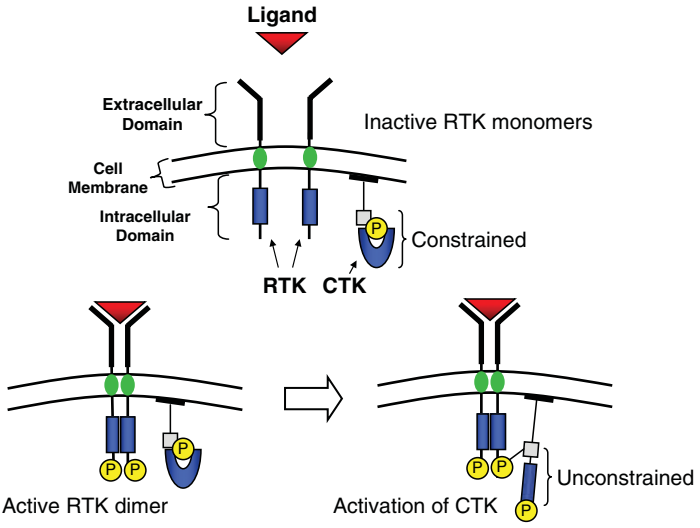

Fig. 5.9 Activation of a protein tyrosine kinase by an extracellular ligand. A generic receptor protein tyrosine kinase (RTK) is composed of an extracellular domain (black) that directly interacts with ligands (red), a transmembrane domain (green) and an intracellular domain that contains a conserved catalytic region (blue). A membrane-associated cytoplasmic tyrosine kinase (CTK) is maintained in inactive form by intramolecular constraints that inhibit its catalytic domain. Upon ligand binding, the RTK molecules dimerize, and activate their catalytic domains by autophosphorylation. The intramolecular constraints that keep the CTK inactive are relieved when the SRC-homology domain (gray) preferentially associates with the phosphorylated form of the RTK dimer. Thus activated, RTK and CTK can trigger downstream pathways

spans the lipid bilayer that composes the cell membrane. An evolutionarily conserved catalytic domain, which contains protein tyrosine kinase activity, resides in the intracellular portion of the molecule. The binding of ligands to RTKs causes the formation of dimers or higher-order oligomers. Concomitant with this oligomerization is the activation of protein tyrosine kinase activity and autophosphorylation of each receptor molecule on specific tyrosine residues. The oligomeric, phosphorylated RTK complex can then function to recruit cytoplasmic protein tyrosine kinases.

There are many variations in the generic mode of RTK function described in Fig. 5.9. Some receptors can recognize several related ligands, others are highly specific. Some types of RTKs, such as those in the epidermal growth factor (EGF) receptor family, are activated by the binding of a single ligand molecule. Others, such as the fibroblast growth factor (FGF) family require the simultaneous binding of two different ligands for activation of kinase activity. The ERBB2 protein requires no ligand at all, but rather interacts with other EGF family members. Within the cytoplasm, some types of RTKs such as those in the FGF receptor family assemble highly ordered complexes of docking proteins that provide an additional level of control.

Several RTK genes are mutated at high frequency in diverse types of cancer. Among the most frequently mutated are *EGFR* and *ERBB2* (see Table 5.2), which are members of the EGF receptor gene subfamily.

Table 5.2 RTK genes altered in cancers

Proto-oncogene	Ligand	Oncogenic alteration	Cancers
EGFR (ERBB1)	Epidermal growth factor (EGF)	Point mutation, deletion	Lung, colorectal, and breast carcinoma
	Transforming growth factor β (TGFβ)	Amplification	Glioblastoma
ERBB2 (HER2/neu)	None	Amplification	Breast, ovarian, gastric, cervical, and lung carcinoma
		Point mutation	Neuroblastoma
MET	Hepatocyte growth factor	Amplification	Medulloblastoma
		Point mutation	Esophageal and gastric carcinoma
			Hereditary papillary renal cell carcinoma
RET	Glial-derived neurotropic factor	Complex rearrangement	Thyroid carcinoma
		Point mutation	Multiple endocrine neoplasia syndromes 2A & 2B
C-KIT	Stem cell factor	Point mutation	Acute myeloid leukemia, germ cell tumors
		Amplification	Glioblastoma
FGFR1	Fibroblast growth factor	Point mutation	Glioblastoma
		Translocation	Acute myelogenous leukemia, lymphoma

By several mechanisms, cancer-associated genetic alterations result in the dysregulation of RTK catalytic activity. The most common genetic alterations that affect RTK genes are point mutations and amplification of the entire gene. Single nucleotide substitutions affecting the extracellular or transmembrane domains can promote receptor dimerization in the absence of ligand (see Fig. 5.10). Alternatively, single nucleotide substitutions can affect a motif within the catalytic domain known as the *activation loop* and cause an increase in basal kinase activity. A third type of genetic alteration that can dysregulate RTK activity is translocation. In cases where the translocated protein domain is normally involved in protein–protein interactions, the mutant receptors dimerize. Translocations are best described in leukemias and lymphomas. In other cancers, amplification of RTK genes can cause receptor overexpression and overactive signaling (see Fig. 5.11).

Once activated, RTKs recruit and activate various signaling molecules in the cytoplasm, including CTKs. CTKs are sometimes referred to as *nonreceptor tyrosine kinases*. CTKs are typically associated with the plasma membrane via an N-terminal posttranslational modification, but they are not transmembrane proteins.

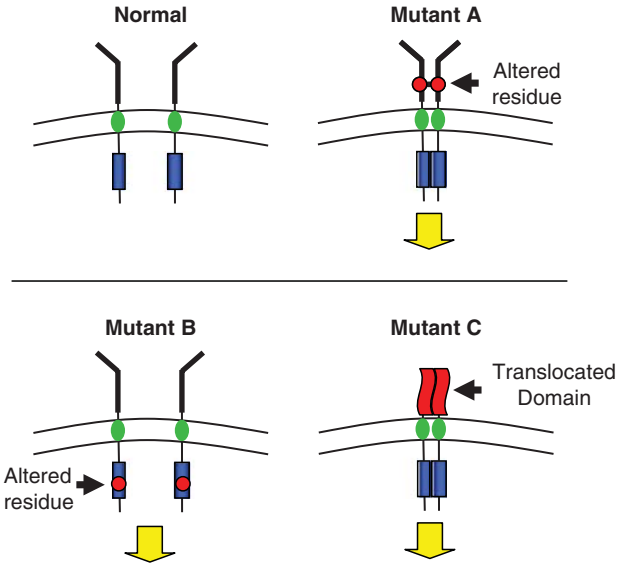

Fig. 5.10 Point mutations can result in RTK dysregulation. Mutant A contains an amino acid substitution (shown in red) in the extracellular domain that causes RTK molecules to have an increased affinity for one another and to dimerize. Mutations in the transmembrane domain can have a similar effect (not shown). Mutant B carries an amino acid substitution mutation in the activation loop of the catalytic domain, increasing the basal kinase activity of RTK monomers. Mutant C is a fusion protein in which the extracellular domain derived from an unrelated protein that is normally 'sticky' and therefore participates in protein–protein interactions. In all cases, signaling is ligand-independent

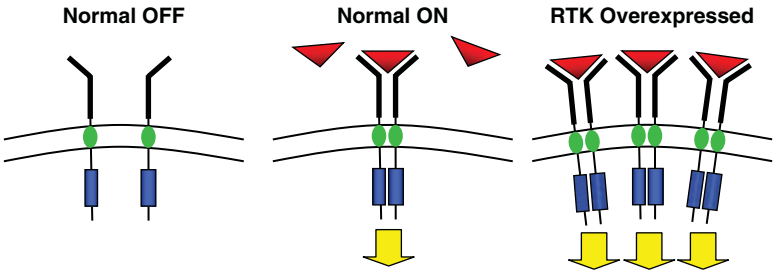

Fig. 5.11 Amplification of RTK genes can cause cells to become hypersensitive to ligand. RTK proteins encoded by wild type genes normally trigger downstream responses (yellow arrow) that depend upon the presence of ligand (red). Amplification of RTK genes leads to overexpression of RTK receptors, and their increased numbers at the cell surface. Although each receptor is normal in structure and function, cells are hypersensitive to ligand

The prototype CTK is C-SRC, encoded by the first identified cancer gene. As described in Chapter 2, the C-SRC protein is normally phosphorylated on a C-terminal tyrosine residue. An N-terminal domain has a high affinity for the phosphorylated C-terminus. C-SRC is thus maintained in a constrained, inactive form. Following an upstream stimulus, the intracellular domains of neighboring RTK molecules become phosphorylated on multiple tyrosine residues (see Fig. 5.9). The N-terminal domain of C-SRC has a greater affinity for the newly phosphorylated RTKs than for its own C-terminus. As a result, C-SRC undergoes a conformational change that results in activation of its kinase domain.

Several oncogene products function in a manner similar to C-SRC and are considered members of a protein family. Proteins in the SRC family exhibit significant amino acid sequence homology. Several regions are critical to their function in signaling, and are known as SRC homology domains SH1, SH2 and SH3. SH1 contains the protein kinase domain. SH2 is required for the binding of the C-terminal phosphotyrosine residue, and SH3 is required for additional protein–protein interactions.

Mutations in *C-SRC* are found at low frequency in several cancers including colorectal carcinomas, in which truncating mutations have been reported. While activated *C-SRC* is not a prevalent oncogene, *C-SRC* and other members of the SRC family are central components of cancer gene pathways that involve RTKs. It is most likely this attribute that led *C-SRC* to be appropriated by the Rous sarcoma virus. It is interesting that *C-SRC* is not found to be mutated nearly as frequently as many other proto-oncogenes. One might infer that in most tissues, there are detrimental effects of *C-SRC* gain of function that must outweigh any potential increases in proliferation during clonal evolution.

The tyrosine kinase encoded by the *ABL* proto-oncogene represents a distinct CTK family. Unlike other CTKs, *ABL* is located in both the cytoplasm and the nucleus. The translocation that generates the *BCR-ABL* fusion gene in chronic

myelongenous leukemia has three effects on ABL signaling. First, because the *BCR* gene promoter is highly active, *BCR-ABL* is overexpressed. Second, the BCR domain of the fusion protein facilitates the formation of oligomeric complexes, which result in the constitutive activation of ABL tyrosine kinase activity. Third, the BCR-ABL fusion protein is excluded from the nucleus, which restricts access to a subset of the substrates normally phosphorylated by ABL.

Membrane-Associated GTPases: The RAS Pathway

A distinct mechanism by which signals at the cell membrane are processed and disseminated involves the binding and hydrolysis of guanosine triphosphate (GTP). Membrane-associated enzymes known as GTPases are activated when GTP is non-covalently bound and inactivated when GTP is hydrolyzed to guanosine diphosphate (GDP). The role of GTPases in cancer cell signaling was revealed by the discovery of the RAS family of oncogenes. RAS-encoded proteins are membrane-bound GTPases that, when present in mutant form, play an important role in the aberrant growth properties of cancer cells.

The three human RAS genes (*H-RAS, N-RAS* and *K-RAS*) encode highly related p21 kilodalton proteins that are localized to the interior surface of the cell membrane. These proteins essentially function as molecular switches that can be turned on and off by the binding and subsequent hydrolysis of GTP, respectively. This binary mode of signaling is highly regulated by additional proteins that regulate a *GDP/GTP cycle* (see Fig. 5.12).

RAS proteins exist in equilibrium between GTP bound (active) and unbound (inactive) states. RAS proteins have intrinsically low levels of GTP binding and hydrolytic activities. GDP/GTP cycling by RAS is modulated by two types of regulatory proteins. *Guanine nucleotide exchange factors* (GEFs) promote formation of the active, GTP-bound state. Two examples of GEFs are the SOS proteins SOS1

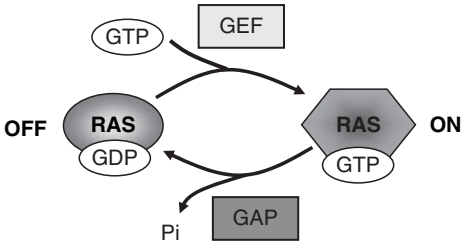

Fig. 5.12 Regulation of RAS-mediated GTP binding and GTP hydrolysis. RAS proteins have low intrinsic GTP-binding activity, which can be greatly stimulated by physical association with a guanine nucleotide exchange factor (GEF). Similarly, the hydrolysis of GTP by RAS is stimulated by a GTPase activating proteins (GAP). Thus the binary mode of signaling of RAS (ON and OFF) is highly regulated

and SOS2, named for the *Drosophila* genes *Sons of Sevenless*, with which they share significant homology. SOS proteins directly interact with RAS and stimulate the exchange of GDP for GTP. The hydrolysis of GTP is facilitated by the interactions of RAS with *GTPase activating proteins* (GAPs). These include the GTPase activating protein known as p120 GAP, and neurofibromin, which is encoded by the *NF1* tumor suppressor gene (see Chapter 3). The GEFs and GAPs control RAS signaling by influencing the balance of GTP-bound and unbound forms.

Membrane association is a requirement for RAS activity. At this location, RAS is activated in response to numerous upstream signals, including those that are emitted by nearby RTKs (see Fig. 5.13). The best characterized RTK-RAS relationship is the one involving the epidermal growth factor receptor (EGFR). Activated by the presence of ligand, the EGFR receptor phosphorylates its own cytoplasmic domain on specific tyrosine residues. An adaptor protein known as GRB2 associates with the activated EGFR complex via an SH2 domain. GRB2 recruits an SOS protein, and thus brings it into the proximity of RAS. SOS stimulates the exchange of GDP for GTP by RAS, triggering a conformational change. Thus activated, RAS can productively interact with downstream molecules. The GAP proteins that function to deactivate RAS also contain a SH2 domain that allows them to be recruited to the membrane.

The conformational change that accompanies the binding of RAS with GTP allows RAS to activate downstream signaling molecules. The most important of these appears to be the RAF family of serine/threonine kinases. In a complex, multi-step process, the three RAF kinases are activated upon their recruitment to the cell membrane by activated RAS. Once activated, RAF proteins phosphorylate and activate two signaling molecules belonging to the MEK family, alternatively known as the MAP kinase kinase (MAPKK) family. The MEKs are unusual protein kinases that have dual specificity; they can phosphorylate proteins on serine/threonine and on tyrosine residues. Activated MEKs translocate to the nucleus and

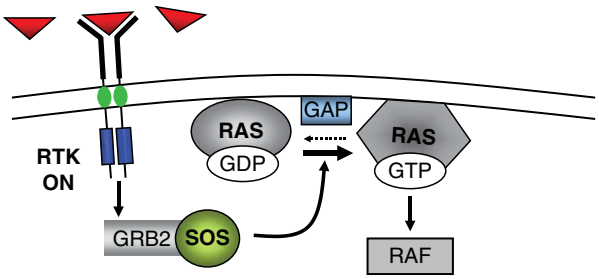

Fig. 5.13 RAS proteins are activated by RTKs. A transmembrane RTK such as EGFR is activated by ligand (red) binding and recruits GRB2 and SOS proteins. SOS is a GEF that triggers the exchange of GDP for GTP and thereby results in a change in RAS conformation. This activated form of RAS can interact with downstream molecules, including the RAF serine/threonine kinases. The GAP proteins, which stimulate GTP hydrolysis, can also be recruited to the cell membrane and attenuate downstream responses

phosphorylate several downstream targets. Among these are two extracellular signal regulated kinases (ERKs), alternatively known as mitogen-activated protein kinases (MAP kinases). Activated ERKs are able to translocate across the nuclear membrane. Thus, via sequential activation of the MEKs and the ERKs, RAF proteins trigger a cascade of protein kinase signaling that spans the cytoplasm and reaches effectors in the nucleus (see Fig. 5.14).

The ERKs are the primary effector of numerous types of proliferative signals. Via phosphorylation, ERKs can directly activate transcription factors and thus affect gene expression. Other important substrates of the ERKs include the ribosomal protein S6 kinases (RSKs), that are regulators of protein synthesis, and the Rho-like GTPases, that have been shown to stimulate changes in cell shape and motility.

The RAS pathway affects many aspects of cell growth that are aberrant in cancer cells. A large body of experimental evidence supports the role of the MEKs and the ERKs as the major effectors of these responses. Inhibition of the RAF, MEK or ERK proteins by the overexpression of dominant inhibitory mutants has been shown to impair the ability of RAS to transform primary cells. Conversely, the overexpression of RAF or MEK can phenocopy the transformation and tumorigenic properties of RAS mutants.

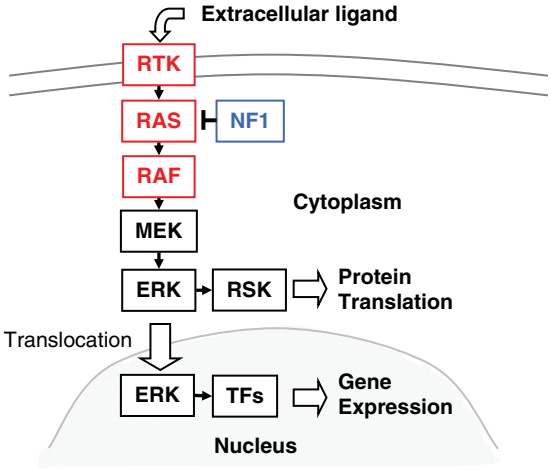

Fig. 5.14 RAS signaling connects RTKs with kinase cascades that alter gene expression and protein translation. RAS represents a node at which upstream and downstream pathways converge. In response to RTK signaling, RAS proteins activate RAF family members. RAS can be deactivated by the GAP proteins, which include the product of the *NF1* gene. RAF proteins phosphorylate and activate the MEKs, which in turn phosphorylate and activate the ERKs. The ERK proteins can activate the ribosome-associated RSK proteins, thereby affecting protein synthesis. ERKs can also translocate into the nucleus and regulate numerous transcription factors (TFs). Thus, RAS signals are transmitted throughout the cell. Proteins that can be constitutively activated via oncogenic mutations are shown in red. NF1 is the product of a tumor suppressor gene and is shown in blue

Genetic Alterations of the RAS Pathway in Cancer

There are several types of genetic alterations that affect RAS signaling in cancers. By far the most prevalent of these are mutations of RAS genes themselves. About 15% of all human cancers contain mutations in a member of the RAS gene family. As described in Chapter 2, the most frequently mutated RAS family member is *K-RAS*. The majority of tumor-associated mutations in *K-RAS* affect codons 12, 13, 59 or 61. While these codons are in the proximity of the guanine nucleotide-binding sites, the effect of the common tumor-associated mutations is not the reduction of intrinsic GTPase activity. Rather, mutant RAS proteins fail to respond to the stimulatory effects of GAP. As a result, mutant RAS proteins remain in the GTP-bound state and are constitutively active.

The tumor suppressor *NF1* encodes a GAP protein. In cells with biallelically inactivated *NF1*, RAS is maintained in its constitutively active, GTP-bound form. Thus, mutations in *NF1* have a similar effect on RAS activity as mutations in RAS genes. As described in Chapter 3, *NF1* mutations are found in both sporadic tumors and in the germline of patients with neurofibromatosis type 1, a syndrome that predisposes affected individuals to cancer.

More widespread than inactivating *NF1* mutations are activating mutations in the RAF family member *BRAF*. The most common *BRAF* mutation is a T→A transversion that changes the normal valine at codon 599 to a glutamic acid residue (V599E). This mutation is located in highly conserved protein kinase motif that is involved in ATP binding. The V599E mutation causes a tenfold increase in basal BRAF kinase activity. Thus, single nucleotide substitutions in *BRAF* can cause constitutive activation of the RAS pathway.

BRAF mutations have been found in the majority of melanomas, and in many colorectal cancers, gliomas, and sarcomas. In melanomas, the mutations in *BRAF* are notably distinct from the UV signature mutations that are frequently associated with skin cancers. In colorectal cancers, the pattern of *BRAF* mutations has been found to reflect the mechanism of genetic instability present in individual tumors. *BRAF* mutations are largely restricted to tumors exhibiting mismatch repair deficiency, while *K-RAS* mutations are found in the majority of colorectal tumors that are mismatch repair proficient (see Chapter 4). Notably, coincident *K-RAS* and *BRAF* mutations do not occur in a single tumor. Studies of this kind provide important evidence that *K-RAS* mutations and *BRAF* mutations are equivalent in their downstream effects.

Mutations in the RAS pathway illustrate key principles of cancer genetics. While RAS mutations are not sufficient nor required for the development of tumors, the inactivation of RAS and downstream components of its signaling pathway are widespread in diverse cancers. In many tissues, there is evidently a significant amount of selective pressure to inactivate the RAS signaling pathway. This is most obvious in highly studied colorectal cancers. In the mucosae of the colorectum, the constitutive activation of RAS signaling occurs as tumors progress from small to intermediately sized adenomas (see Chapter 2). In most cases, stimulus-independent RAS signaling can be attributed to the mutation of RAS; approximately half of all

colorectal tumors accumulate an activating mutation in a RAS gene. In a significant fraction of cancers, other mutations arise that have a similar effect as RAS mutations. For example, mismatch repair-deficient colorectal tumors frequently harbor a mutated *BRAF* allele. Once the RAS pathway is constitutively turned on via genetic alteration, there is no remaining selective pressure to inactivate additional genes that populate the pathway.

Membrane-Associated Lipid Phosphorylation: The PI3K/AKT Pathway

In response to mitogenic ligands, receptor tyrosine kinases (RTKs) can trigger the activation of a class of enzymes known as phosphatidylinositol 3-kinases (PI3Ks). This unique class of enzymes catalyzes the phosphorylation of inositol-containing lipids. These phospholipids then act as second messengers that stimulate downstream signaling molecules. PI3K activation represents a distinct pathway that is triggered by receptor tyrosine kinase (RTK) signaling.

PI3K was first discovered in the 1980s as a novel enzymatic activity associated with partially purified viral oncoproteins such as C-SRC. Further purification and analysis revealed that this activity could be attributed to a two-subunit enzyme complex. An 85 kDa protein (p85) associates with RTK proteins and serves a regulatory function. Catalytic activity of PI3K heterodimers is contained in a separate 110 kDa protein (p110). The isolation of the genes that encode the catalytic and regulatory domains of PI3K has revealed a large and complex family of proteins. The PI3K proteins known to be involved in cancer belong to a subcategory known as Class 1A.

Phosphatidylinositol phosphates contain a fatty acid moiety that is associated with the inner surface of the cell membrane. A glycerol backbone links the fatty acid moiety to a inositol head group that is the target of both lipid kinases and lipid phosphatases (see Fig. 5.15). Class 1A PI3Ks catalyze the phosphorylation of phosphatidylinositol (4,5) bisphosphate (called PIP_2) to produce phosphatidylinositol (3,4,5) triphosphate (called PIP_3). The catalytic and the regulatory domains of the Class 1A PI3Ks are each encoded by a family of three distinct genes. The activating subunits contain SH2 and SH3 domains that interact with RTKs. The catalytic subunits contain protein domains that are critical for kinase activity, membrane anchoring and interactions with the regulatory subunit.

Several RTKs are known to trigger PI3K dependent signaling. Among these are the well-characterized EGF receptor (EGFR) and receptors that respond to the growth stimulatory effects of insulin.

Normal cells that are unstimulated by mitogenic ligands have very low levels of PIP_3. Following ligand-dependent RTK activation and autophosphorylation, a PI3K complex (composed of a p85 and a p110 subunit) is recruited to the receptor by the SH2 domain of p85. Prior to its activation, p85 exerts and inhibitory effect on p110. The RTK-p85 interaction relieves this inhibition, and also brings p110 in close

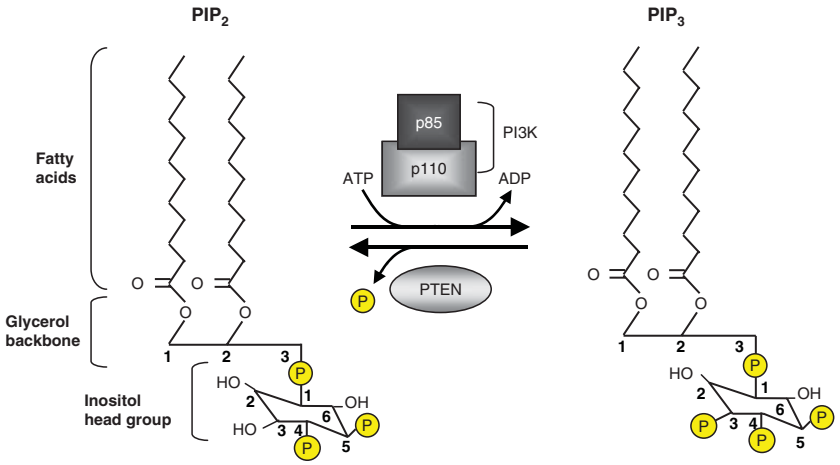

Fig. 5.15 Regulation of the PIP_2–PIP_3 cycle by PI3K and PTEN. The heterodimeric PI3K complex catalyzes the ATP-dependent phosphorylation of phosphatidylinositol (4,5) bisphosphate (PIP_2) at the D3 position of the inositol moiety to generate phosphatidylinositol (3,4,5) triphosphate (PIP_3). The reverse reaction is catalyzed by the lipid phosphatase encoded by the *PTEN* tumor suppressor gene. Relevant phosphate groups are shown in yellow

proximity to its lipid substrates at the cell membrane. Levels of intracellular PIP_3 increase as a result of the increased catalytic activity of p110.

Newly generated PIP_3 acts as a second messenger that activates downstream signaling proteins known as the AKT family. The three closely related AKT proteins are encoded by the cellular homologs of the viral oncogene *V-AKT*, originally isolated from a mouse thymus tumor. AKT proteins binds PIP_3 via a protein domain originally defined in the cytoskeletal protein *pleckstrin*. As a result of the pleckstrin homology domain-PIP3 interaction, AKT is recruited to the inner surface of the cell membrane, where it is phosphorylated and activated by the phosphoinositide-dependent protein kinases (PDKs).

AKT proteins are serine/threonine protein kinases that are involved in multiple downstream pathways that control cell growth, proliferation, motility and death (see Fig. 5.16):

Promotion of cell cycle progression. AKT inhibits negative regulators of cell cycle progression. Among these is p21, the product of the *CDKN1A* gene. High levels of p21 inhibit the cyclin-dependent kinases that promote the progression of the cell cycle. Phosphorylation of p21 by AKT causes the sequestration of p21 in the cytoplasm, thereby preventing it from exerting its regulatory effect in the nucleus. The inhibition by AKT of glycogen synthetase 3 (GSK3) is another mechanism by which AKT regulates cyclin-dependent kinase activity. Cyclin D is directly inhibited by GSK3; this inhibition is relieved by AKT-dependent inhibition of GSK3.

Downregulation of apoptosis. AKT directly phosphorylates several substrates that are directly involved in apoptosis. Among these are BAD, a proapoptotic

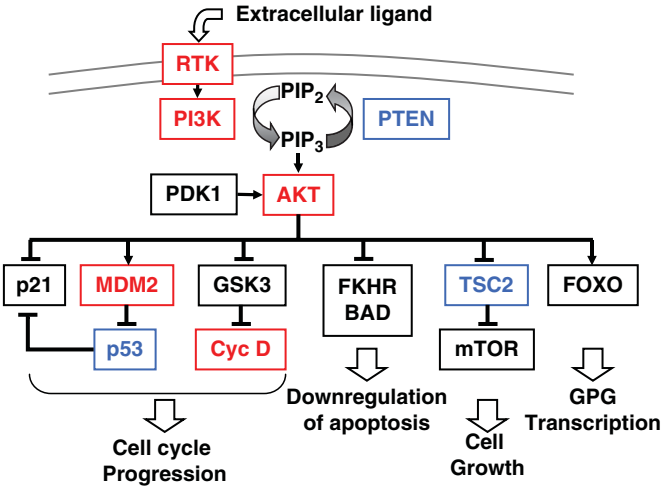

Fig. 5.16 The PI3K/AKT pathway. Ligand-dependent activation of RTK signaling causes the activation of PI3K, and the generation of PIP_3. Via its pleckstrin homology domain, AKT binds PIP_3 and is thus recruited to the inner surface of the cell membrane. AKT is activated by a dual regulatory mechanism that requires translocation and subsequent phosphorylation by PDK1. Active AKT phosphorylates numerous downstream substrates; only the representative ones are shown. Cell cycle progression is stimulated by the AKT-dependent phosphorylation of the cyclin-dependent kinase inhibitor p21. Expression of p21 is also inhibited by the MDM2-dependent inhibition of p53. The activity of cyclin D is increased by the AKT dependent inhibition of glycogen synthetase kinase 3B. Apoptosis is downregulated by inhibitory signaling to several proapoptotic proteins. AKT inhibits the mTOR pathway via inhibition of TSC2, and thereby promotes protein biosynthesis. The expression of growth-promoting genes is increased by the activation of the FOXO family of transcription factors. Proteins encoded by proto-oncogenes are shown in red; tumor suppressor gene products are shown in blue

member of the BCL2 family that is inhibited by AKT-dependent phosphorylation. AKT also blocks the activity of the transcription factor FKHR, which induces expression of pro-apoptotic genes.

Control of gene transcription. In addition to the regulation of FKHR, AKT controls gene expression by several mechanisms. Direct activation of a family of transcription factors known as FOXO causes the increase in the expression of growth promoting genes, while activation of MDM2 antagonizes p53 and thus prevents the expression of growth inhibitory genes.

Control of the mTOR pathway. An important regulator of cell growth is the pathway controlled by the mammalian target of rapamycin (mTOR). mTOR is a serine/threonine kinase that regulates protein synthesis on the basis of the availability of nutrients. The downstream effect of mTOR activation is an increase in protein translation, a key requirement for cell growth. AKT regulates mTOR via the negative regulation of TSC2. TSC2, along with TSC1, inhibits mTOR. By inhibiting TSC2, AKT increases the activity of mTOR and thereby promotes protein biosynthesis.

Genetic Alterations of the PI3K/AKT Pathway in Cancer

The importance of PI3K signaling is suggested by the many cancers in which it is disrupted by mutations. During tumorigenesis, the PI3K pathway is frequently affected by the activation of proto-oncogenes and by the inactivation of tumor suppressor genes. The mutations of AKT regulators result in constitutive pathway activation.

PIK3CA encodes a p110 subunit of the PI3K heterodimer. As described in Chapter 2, *PIK3CA* is a proto-oncogene that is frequently mutated in many common types of cancer, including colorectal, breast, brain and ovarian cancers. *PIK3CA* is activated primarily by point mutations in one of two hotspots that determine overall enzyme structure and lipid kinase activity (see Fig. 5.17). In addition, amplification of *PIK3CA* is seen in some cancer types, most prominently in ovarian carcinomas. Oncogenic mutations and amplification increase PIK3CA kinase activity and trigger constitutive activation of AKT. Related genes that encoded other p110 family members do not seem to be altered in cancers to a significant extent.

The phosphorylation of PIP_2 by PI3K is antagonized by the lipid phosphatase encoded by *PTEN* (see Fig. 5.15). As described in Chapter 3, *PTEN* is a tumor suppressor gene that is frequently inactivated in many sporadic cancers. Additionally, germline mutations of *PTEN* are a cause of inherited cancer predisposition. Although PTEN can function as a protein kinase, it appears that its tumor suppressor function is largely linked to its role in the dephosphorylation of PIP_3.

Amplification involving the AKT family member *AKT2* has been observed in diverse cancers, including those of the pancreas, ovary, uterine cervix, the head and neck and lung. Interestingly, amplification of both *AKT2* and *PIK3CA* has been shown to coexist in some ovarian tumors. Amplification of these two genes in a single tumor would seem contrary to the idea that mutations within a pathway tend to be mutually exclusive. There is abundant evidence for such exclusivity in other

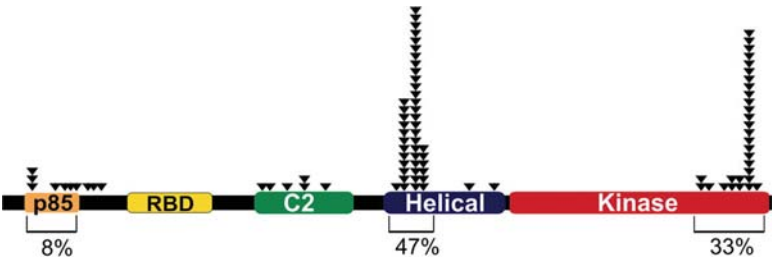

Fig. 5.17 Mutations of *PIK3CA* in colorectal cancers. The protein coding region of *PIK3CA* contains multiple functional domains. These include a binding site for the PI3K regulatory subunit (p85), a putative binding domain for the RAS oncoprotein (RBD), a membrane anchoring domain (C2), a helical domain and a lipid kinase domain. Arrowheads indicate the positions of missense mutations. The percentage of total *PIK3CA* mutations that occur within each domain is indicated. (Reprinted from Samuels et al. 2005. With permission from AAAS.)

pathways, primarily the dearth of coincident mutations in *MDM2/P53* and *K-RAS/BRAF*. The violation of this principle in the PI3K/AKT pathway is probably the result of the subtle phenotypic effects caused by gene amplification. Both AKT2 and PIK3CA require activation by upstream enzymes. In both cases, amplification of the wild type gene would be predicted to result in only a modest increase in activity. In contrast, activating point mutations of *PIK3CA*, as well as in genes such as *K-RAS*, lead to large increases in enzymatic activity that serves to fully activate downstream responses. In the case of AKT2, it appears that the small increase in AKT signaling that results from *AKT2* amplification can be augmented by coincident amplification of *PIK3CA*. The relatively small phenotypic effect of gene amplification might provide the selective pressure to amplify both genes during tumorigenesis. Notably, the coexistence of *AKT2* amplification and *PIK3CA* activation by point mutation has not been observed.

Morphogenesis and Cancer: The WNT/APC Pathway

Communication between neighboring cells is vital for the normal development of multicellular organisms. In adult humans, mutations that affect cell to cell communication can cause a loss of tissue homeostasis that leads to cancer. A signaling pathway that is important in embryonic development and is also dysregulated in some types of cancer is triggered by ligands known as *WNTs*. WNTs are secreted proteins that become soluble and active upon the covalent attachment of a lipid moiety. In their lipid-modified form, WNTs function as ligands that trigger evolutionarily conserved morphogenic pathways. The WNTs comprise a large family of proteins that play critical roles during mammalian development. However, it is the function of WNT signaling in adult tissue homeostasis that is directly relevant to tumorigenesis.

Insights into the role of WNT signaling and the structure of the downstream WNT-dependent pathways were first gained from studies of development in *Drosophila*. A ligand required for wing development, designated Wg, activates a pathway that mediates fundamental developmental processes that include embryonic induction, generation of cell polarity and the specification of cell fate.

The mammalian homologs of Wg were originally identified as genes that were called INTs because they were found near the sites of murine retrovirus *in*sert*i*on. Tumors were found to form in mice in which retrovirus integration caused INT genes to become overexpressed. Significant sequence homology between Wg and INT led to the redesignation of the mammalian genes as WNTs. These evolutionarily conserved ligands reveal a link between altered developmental processes and cancer.

There are several distinct pathways that are triggered by the large family of WNT ligands. The WNT-dependent pathway that is populated by cancer genes is alternatively referred to as the *canonical* WNT pathway or the WNT/APC pathway (see Fig. 5.18). The upstream receptor of the WNT/APC pathway is a

transmembrane protein known as Frizzled, which like other proteins in the pathway is named after its *Drosphila* homolog. Frizzled has WNT-binding domains on its extracellular domain and seven transmembrane domains. Unlike receptor tyrosine kinases, frizzled does not appear to have ligand-specific enzymatic activity. Rather, upon ligand-binding Frizzled associates with a second type of transmembrane lipid-modified protein called the liporeceptor-related protein (LRP). It is believed that LRP functions as a co-receptor complex with Frizzled and activates an intracellular protein called Dishevelled.

An important consequence of WNT/APC pathway activation is the stabilization of a protein known as β-catenin. In the absence of upstream signals, β-catenin is phosphorylated by the protein kinase GSK3 (see Fig. 5.18). Phosphorylation of β-catenin promotes the formation of a physical complex containing the tumor suppressor protein APC and a protein called AXIN. This complex functions in the ubiquitin-dependent degradation of β-catenin by the proteosome. When WNT signaling is triggered by ligand binding, the activation of Dishevelled by the membrane-associated signaling complex leads to the inhibition of GSK3 kinase activity. As a result, β-catenin dissociates from the degradation complex, is stabilized, and thus can translocate to the cell nucleus. Nuclear β-catenin associates with and activates the T-cell factor (TCF)-family of transcription factors. The growth

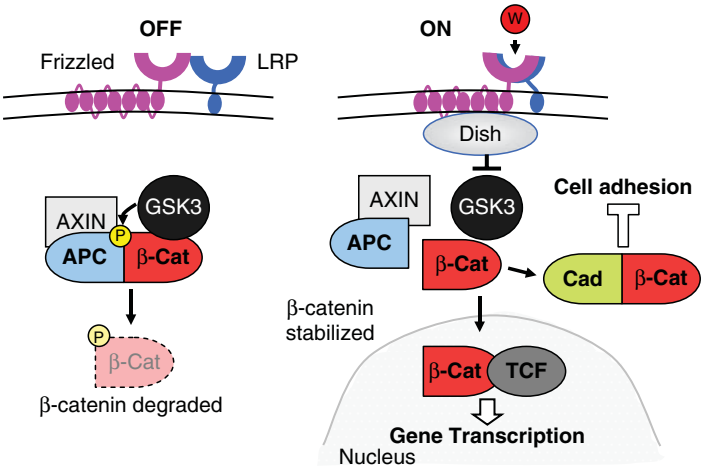

Fig. 5.18 The canonical WNT signaling pathway. In the absence of WNT ligand (OFF; left panel), phosphorylation of β-catenin by GSK3 favors the formation of a complex composed of APC and AXIN. β-catenin is targeted for degradation when the WNT pathway is OFF. When the pathway is turned on by ligand, Frizzed and LRP cooperatively activate Disheveled at the cell membrane, which functions to inactivate GSK3. In the absence of GSK3-mediated phosphorylation, the degradation complex is dissociated and β-Catenin is stabilized. β-catenin then translocates to the nucleus and, in cooperation with the TCF family of transcription factors, activates the expression of growth promoting genes. In addition, cytoplasmic β-catenin can associate with the cadherin proteins, which mediate cell adhesion. The binding of β-catenin to APC or to cadherin proteins is mutually exclusive

promoting targets of β-catenin-activated TCF transcription include the transcription factor C-MYC and the cell cycle regulator Cyclin D.

β-catenin is a multifunctional protein that is involved in multiple cellular processes. In addition to the TCF transcription factors in the nucleus, β-catenin also associates with cytoplasmic proteins known as *cadherins*. The cadherins are calcium-dependent proteins that function to mediate adhesion between cells and their surrounding matrix of structural proteins. The interaction between β-catenin and the cadherins suggest a role for WNT/APC signaling in the regulation of cellular organization. This relationship remains incompletely understood.

Inactivation of the WNT/APC Pathway in Cancers

Inactivating mutations in *APC* occur frequently in colorectal cancers. As described in Chapter 3, germline mutations in *APC* cause polyposis and confer a markedly increased lifetime risk of cancer, while somatic mutations of *APC* frequently underlie sporadic tumors. Even subtle *APC* mutations not predicted to completely inactivate the encoded protein apparently lead to an increase in polyp formation, indicating the great extent to which APC functions as a gatekeeper in the colorectal mucosa. The role of the canonical WNT pathway in colorectal tumorigenesis is strongly supported by the central regulatory role of APC in this pathway, and by the observation that colorectal cancer cells typically exhibit upregulated TCF activity.

APC is inactivated in most, but not all, colorectal cancers by loss of both wild type alleles. However, some colorectal cancers retain wild type *APC* alleles. In the majority of such cancers, mutations in other genes disrupt canonical WNT signaling. The elucidation of the canonical WNT pathway demonstrated the regulatory role played by several proteins, including β-catenin and AXIN. Large-scale DNA sequencing efforts have revealed that mutations in the genes that encode these proteins contribute to a small fraction of colorectal cancers.

β-catenin, encoded by the *CTNNB1* gene, is phosphorylated by GSK3 kinase on several N-terminal serine and threonine residues. Phosphorylation of these residues enhances the binding of β-catenin to APC. *CTNNB1* is therefore a proto-oncogene that can be activated by mutations that affect the association between β-catenin and APC. Point mutations or small deletions that alter *CTNNB1* codons 30, 33, 37 and 45, which encode the GSK3 phosphorylation sites, have been found in cancers with wild type *APC*. The form of β-catenin encoded by cancer-associated mutant *CTNNB1* alleles exhibits reduced APC binding. Cells that harbor oncogenic *CTNNB1* exhibit WNT ligand-independent stabilization of β-catenin and increased activation of TCF-mediated transcription. Thus, the mutant form of β-catenin causes constitutive activation of the canonical WNT pathway.

Mutations in *CTNNB1* and *APC* are mutually exclusive. As loss of APC or mutation of the GSK3 phosphorylation sites on β-catenin each affect the binding of APC to β-catenin, it is reasonable to conclude that the disruption of this binding event can be achieved equally by mutational inactivation of *APC* or activation of *CTNNB1*. This is another illustration of an important principle of cancer genetics.

The activation of a pathway by one type of mutation lowers the selective pressure for additional mutations that affect the same pathway in the same manner. A similar relationship exists between *K-RAS* and *BRAF*, which populate the RAS pathway.

In addition to APC, AXIN functions as a negative regulator of the canonical WNT signaling pathway. The two genes that encode AXIN proteins, *AXIN1* and *AXIN2*, are mutated in a variety of cancers. Truncating mutations in these genes have been found that would be predicted to completely inactivate AXIN proteins. Such mutations are sufficient to inhibit the binding of the GSK3 kinase and cause the stabilization of the β-catenin degradation complex. Missense mutations have also been reported, though the functional consequences of such mutants remain incompletely understood. By both genetic and functional criteria, mutations in AXIN genes are not equivalent to mutations in either β-catenin or APC.

Mutations in the gene that encodes the transcription factor TCF4, one of the transcription factors that bind stabilized β-catenin, have been found in many colorectal cancers that feature microsatellite instability (MSI). Cancer-associated TCF4 mutations are frameshifts in a poly-A tract in the C-terminal exon. Such mutations are known to be caused by mismatch repair deficiency (see Chapter 4); their affect on TCF-dependent transcriptional transactivation remains unclear.

Colorectal cancers feature several types of genetic alterations that affect the canonical WNT signaling pathway (see Fig. 5.19). These alterations are also found

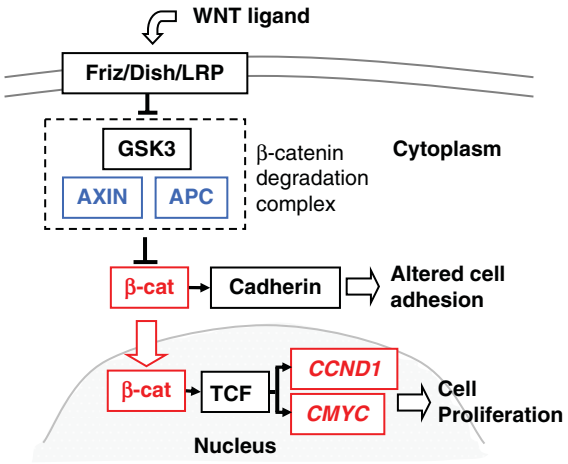

Fig. 5.19 Constitutive WNT signaling in cancer. Several types of mutations can mimic the effects of activated WNT signaling. Mutations in *APC* in colorectal cancers frequently disrupt the β-catenin degradation complex, leading to WNT ligand-independent stabilization of β-catenin. Oncogenic mutations that affect the APC-binding site of β-catenin have the same effect. Inactivating mutations in AXIN-encoding genes can also disrupt the β-catenin degradation complex. Stabilized β-catenin alters cell adhesion via interactions with cytoplasmic cadherins, and promotes transcriptional transactivation by TCFs. Among the targets of β-catenin/TCF transcription are the proto-oncogenes *CCND1*, which encodes the cell cycle regulator cyclin D, and *C-MYC*. Proteins and genes affected by oncogenic mutations are shown in red, tumor suppressors are shown in blue

at lower frequency in other types of cancer. The biallelic inactivation of *APC*, which is characteristically found in colorectal cancers, occurs in a much smaller proportion of sporadic lung, ovarian and breast cancers. Germline mutations of APC that cause the development of colorectal cancer also predispose carriers to several other types of cancer as well (see Chapter 3). Conversely, oncogenic mutations in the β-catenin gene *CTNNB1* are more common in cancers outside the gastrointestinal tract. *CTNNB1* mutations occur in up to one half of all ovarian cancers, and have also been reported in melanomas, cancers of the liver and in Wilms' tumor, a pediatric kidney cancer. Inactivating mutations affecting AXIN have been found in liver cancers. These findings demonstrate that the WNT pathway is disrupted in different ways in different types of cancer. In the epithelial tissues of the colorectum, disruption of canonical WNT signaling appears to be an essential, early step of tumorigenesis. In many other types of tissues, there is no apparent evolutionary benefit conferred by loss of APC or WNT signaling.

Interestingly, one critical component of the WNT signaling pathway that does not appear to be affected by cancer-associated mutations is GSK3. GSK3 also functions downstream of the PI3K/AKT pathway (see previous section). The negative regulation of GSK3 kinase activity is thus a downstream response of two convergent pathways that are frequently inactivated in cancer. One might predict that GSK3 itself might be encoded by a tumor suppressor gene.

Why are inactivating mutations affecting GSK3 not found in cancers? There are several possible explanations. The GSK3 family encompasses the related kinases GSK3α and GSK3β. Redundant functions between these two proteins may render this node relatively impervious to complete loss of function, as inactivation of four alleles would be required to achieve a selective advantage. Alternatively or additionally, there may be negative consequences to the loss of GSK3 activity that may outweigh any proliferative benefits of GSK3 loss of function.

TGF-β/SMAD Signaling Maintains Tissue Homeostasis

Developmental processes such as cell proliferation, differentiation, migration and apoptosis are highly regulated by signaling molecules known as cytokines (described earlier in this chapter). A cytokine called transforming growth factor β1 (TGF-β1) is the prototype of a large family of soluble ligands that regulate development and also function to preserve tissue homeostasis in developed tissues. Like the WNT/APC pathway, the TGF-β pathway links defective development to cancer. Genes that populate the signal transduction pathways downstream of TGF-β are frequently altered in several common forms of cancer.

The receptors that interact with TGF-β ligands are transmembrane serine/threonine protein kinases, and therefore enzymatically distinct from the RTKs. There are two general types of TGF-β receptors, known as type I and type II, each comprised of five distinct proteins. Each TGF-β ligand binds to a characteristic combination of type I and type II receptors. A specific TGF-β ligand/receptor

combination triggers what is referred to the canonical TGF-β pathway, which functions in tumor suppression.

As is the case with other types of ligand, signaling by TGF-β family members is initiated by binding of ligand to specific receptors on the surface of the cell. Homodimeric TGF-β proteins induce the assembly of a receptor complex that includes type I and type II receptor proteins and various accessory receptor proteins (see Fig. 5.20). Upon receptor complex assembly, the constitutively active type II receptor is brought into proximity of the type I receptor, which is then phosphorylated. Thus activated, the type I receptor phosphorylates the signaling proteins SMAD2 and SMAD3. These proteins form several different multimeric complexes with SMAD4. SMAD2/3-SMAD4 complexes shuttle into the nucleus via the nuclear pore complex. In the nucleus, SMAD2/3-SMAD4 complexes cooperate with nuclear transcription factors to regulate the expression of a wide variety of genes involved in cell proliferation, differentiation, migration and death.

Despite the apparent simplicity of the TGF-β/SMAD pathway, the canonical TGF-β signaling pathway can respond to numerous ligands and generate diverse cellular responses. In contrast with the best-characterized RTK pathways, the TGF-β pathways appear to be differentially sensitive to varying concentrations of extracellular ligand. In addition, many intracellular proteins not directly required for signaling can promote or inhibit SMAD-complex formation. Many of the

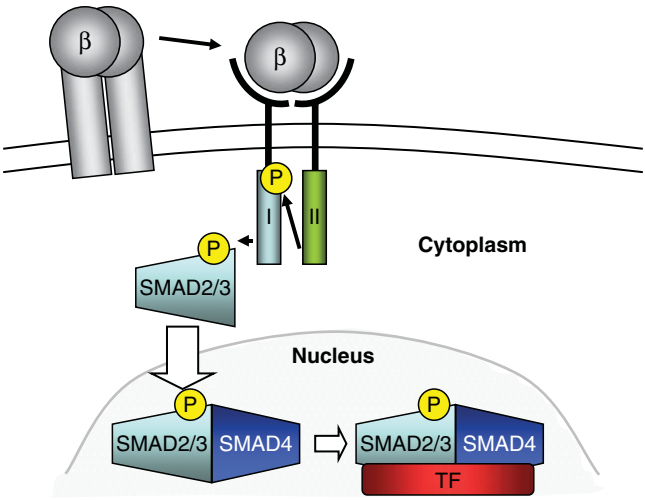

Fig. 5.20 Activation of canonical TGF-β signaling. Accessory receptors (shown in gray) bind homodimeric TGF-β ligand molecules and present them to the specific TGF-β receptors. Upon ligand binding, the type I and type II TGF-β receptors form a complex, causing the phosphorylation of the type I receptor on specific serines and threonines by the type II receptor. The activated type I receptor phosphorylates SMAD2 and SMAD3 proteins in the cytoplasm. SMAD2 and SMAD3 then form multimeric complexes with SMAD4 in the nucleus. These complexes associate with transcription factors and function to both transactivate and repress downstream gene expression

TGF-β responses are cell type-specific. For example, the same ligand might cause one type of cell to proliferate, but trigger a second type of cell to undergo apoptosis. The molecular mechanisms that underlie the wide range of TGF-β responses remain incompletely understood.

Canonical TGF-β signaling suppresses the growth of most normal cells. Several downstream genes that are regulated at the transcriptional level by TGF-β/SMAD pathway activation actively suppress growth (see Fig. 5.21). The genes *CDKN1A* and *CDKN2B* encode the cyclin-dependent kinase inhibitors, p21 and p15, respectively, that function to arrest cell cycle progression. In addition, activated SMAD complexes also appear to repress the transcription of a subset of target genes. These include *C-MYC*, and the inhibitor of differentiation gene *ID1*. Thus, via transcriptional transactivation and transcriptional repression, TGF-β signaling causes normal cells to halt proliferation and to differentiate.

Many types of cancer cells are resistant to canonical TGF-β signaling, and proliferate and remain undifferentiated despite the presence of TGF-β ligand. Several genes that compose the TGF-β/SMAD pathway are targets of inactivating mutations in cancers. The gene that encodes the TGF-βRII receptor is frequently mutated in cancers that are mismatch repair (MMR) deficient. *SMAD4* is a tumor suppressor gene that is inactivated in the majority of pancreatic cancers and also in a significant fraction of colorectal cancers. *SMAD2* is located in close proximity

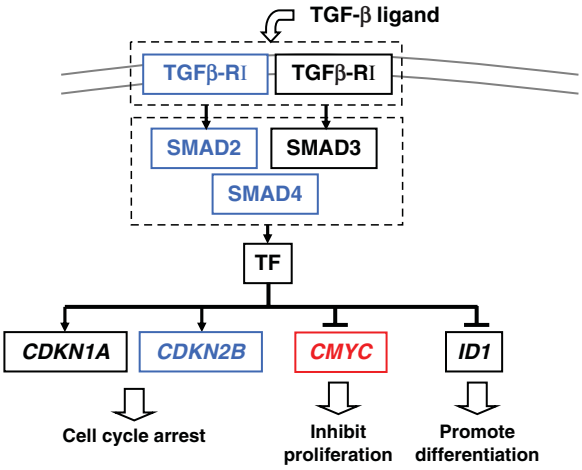

Fig. 5.21 Disruption of the TGF-β/SMAD signaling pathway by tumor suppressor gene mutations. The TGF-β receptor complex can be inactivated via mutations in the gene encoding the type II receptor (TGF-βRII). The SMAD signaling complexes can be disrupted by mutations in either *SMAD2* or *SMAD4*. Due to their close proximity on chromosome 18q, these genes are frequently inactivated by deletions in the same cancers. Normal signaling promotes cell cycle arrest, growth inhibition and differentiation. Inactivation of TGF-β/SMAD signaling in cancers results in a loss of cell cycle arrest mediated by *CDKN1A* and *CDKN2B*, and the promotion of proliferation and loss of differentiation caused by the disinhibition of *C-MYC* and *ID1*, respectively

to *SMAD4* on chromosome 18q, and is therefore lost in many of the large deletions that affect *SMAD4*. In some cancers, *SMAD2* is affected by mutations that do not appear to affect *SMAD4*, indicating that *SMAD2* also functions as a tumor suppressor. Inherited mutations in *SMAD4* cause Juvenile Polyposis Syndrome, a disease that confers cancer predisposition (Chapter 3).

C-MYC is a Downstream Effector of Multiple Cancer Gene Pathways

The *C-MYC* proto-oncogene encodes a transcription factor that regulates the expression of numerous genes in response to both extracellular and intracellular signals. In general, C-MYC induces genes that are required for proliferation and represses genes that are involved in the maintenance of tissue homeostasis. It has been estimated that 15% of all human protein-coding genes are either induced or repressed by C-MYC. The C-MYC target genes are involved in diverse cellular processes including metabolism, cell cycle regulation, apoptosis, protein synthesis, angiogenesis and cell–cell adhesion. These broad cellular effects probably explain why *C-MYC* is frequently activated in many types of cancer. Increased expression of *C-MYC* occurs up to 20% of all human cancers, most often as a result of gene amplification (see Chapter 2). Other MYC family members are also amplified and mutated in cancers, at lower frequencies.

MYC-encoded proteins are transcriptional transactivators that bind specific promoter sequences upstream of target genes. Each of the MYC proteins contains two domains that are highly characteristic of transcription factors. A C-terminal domain known as a *basic helix-loop-helix leucine zipper* (bHLHLZ) facilitates both protein–protein interactions as well as sequence-specific DNA binding. At the N-terminus of the MYC proteins is a transactivation domain that is involved in the recruitment of additional transcription factors.

MYC proteins do not function in monomeric form, but rather complex with one of several factors known as *MYC-associated protein X* (MAX) proteins. The MAX proteins also contain a bHLHLZ motif, which facilitates MYC/MAX heterodimer formation. Partnered with MAX, MYC binds to a defined DNA sequence motif known as an E box. Unlike the short-lived MYC proteins, MAX proteins are constitutively expressed at high levels and are highly stable. Thus, the transcriptional activity of the MYC/MAX heterodimer is limited by the intracellular concentration of MYC protein. The MYC/MAX heterodimer is required target gene transcription but also for repression. The mechanism by which MYC proteins repress the transcription of specific genes is incompletely understood, but involves a third category of proteins known as MADs. Like MYC proteins, MAD proteins form heterodimers with MAX and these appear to compete with MYC/MAX complexes for DNA binding.

In normal cells, *C-MYC* expression is tightly regulated and highly sensitive to upstream signals. Both *C-MYC* transcripts and C-MYC proteins are highly labile

and thus have short half-lives. Because of this lability, increased levels of *C-MYC* transcription can lead to a rapid increase in C-MYC protein activity.

The half-life of C-MYC protein can be affected by upstream signals (see Fig. 5.22). Two RTK-dependent pathways have been shown to increase MYC half-life: the RAS/RAF pathway and the PI3K/AKT pathway. Like many highly regulated proteins, C-MYC is posttranslationally modified. Phosphorylation of a specific residue in the N-terminus, serine 62, has been found to increase C-MYC stability. Serine 62 phosphorylation occurs upon RAS/RAF activation, probably via ERK activation. A second phosphorylation event, on threonine 58 can trigger C-MYC degradation. Phosphorylation at this site is inhibited by PI3K signaling, probably via inhibition of GSK3 kinase. In normal cells, these pathways likely serve to ensure that any activation in C-MYC activity is transient in nature. In cancer cells, C-MYC can be stabilized by enhanced stabilization (by RAS pathway activation) or by suppression of degradation (by PI3K/AKT pathway activation, see Fig. 5.23).

C-MYC is also highly regulated at the level of its transcription (see Fig. 5.24). As described in the previous sections, both the WNT/APC pathway and the TGF-β/

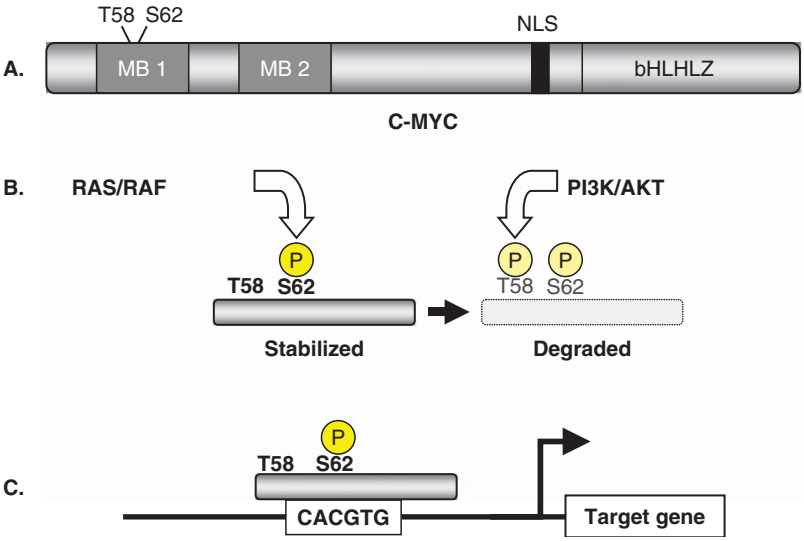

Fig. 5.22 Regulation of C-MYC protein stability. **A.** A schematic representation of C-MYC protein shows the regions of homology shared by all MYC proteins, known as MYC Homology boxes 1 and 2, the nuclear localization signal and the basic helix-loop-helix leucine zipper (bHLHLZ) motif at the C-terminus. **B.** Two residues in MYC homology box 1 have been shown to receive signals from upstream kinases. In response to RAS/RAF signaling, serine 62 (S62) is phosphorylated, resulting in stabilization of C-MYC. Conversely, threonine 58 (T58) is phosphorylated is inhibited by PI3K/AKT signaling. T58 phosphorylation, which occurs after S62 phosphorylation, is associated with C-MYC degradation. Together these events insure that an increase in C-MYC is transient. **C.** Stabilization of C-MYC increases the occupancy of the C-MYC consensus binding site (CACGTG) in the promoter of C-MYC target genes

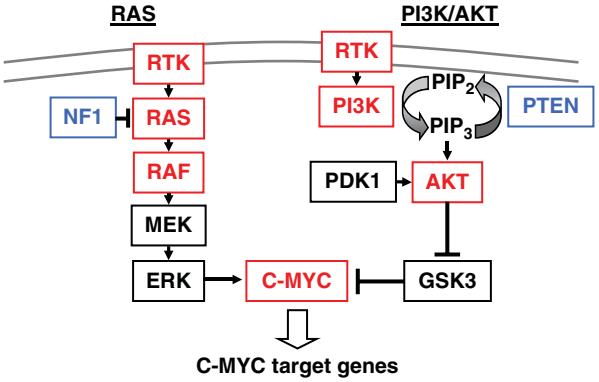

Fig. 5.23 Stability of C-MYC protein is promoted by pathways that are constitutively activated in cancers. RAS signaling causes the activation of ERK, a stabilizer of C-MYC protein. PI3K/AKT signaling causes the inhibition of GSK3, a destabilizer of C-MYC. The activation of either or both of these pathways results in the stabilization of C-MYC protein. Thus, C-MYC protein is a point of convergence of two signaling pathways that are often constitutively activated during tumorigenesis

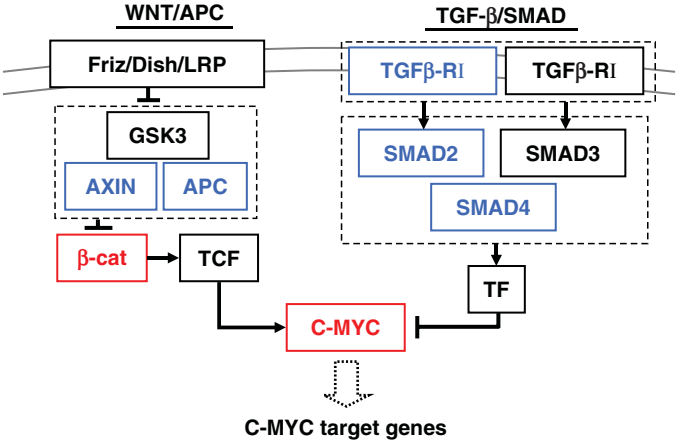

Fig. 5.24 Transcription of *C-MYC* RNA is inhibited by pathways that are inactivated in cancers. WNT/APC signaling inhibits *C-MYC* transcription via the destabilization of β-catenin. In contrast, TGF-β/SMAD signaling activates transcription complexes that repress the transcription of *C-MYC*. Activation of either or both pathways reduces expression of *C-MYC* and also reduces the expression of C-MYC target genes. Conversely, expression of *C-MYC*, and C-MYC target genes, is increased in cancers that harbor mutations that inactivate the WNT/APC and/or the TGF-β/SMAD pathways

SMAD pathways inhibit *C-MYC* transcription. β-catenin promotes the transcriptional transactivation of C-MYC. The WNT/APC pathway antagonizes *C-MYC* transcriptional transactivation by promoting β-catenin degradation. In contrast, activation of the TGF-β/SMAD pathway causes the formation of protein complexes

that transcriptionally repress the *C-MYC* promoter. In cancers with either inactivated WNT/APC or TGF-β/SMAD signaling, *C-MYC* expression is elevated.

In summary, cancers increase C-MYC-dependent transcription in several ways: (1) the *C-MYC* gene itself is transcriptionally induced by constitutively activated cancer gene pathways including the WNT/APC and the TGF-β/SMAD signaling pathways; (2) *C-MYC* is frequently amplified in some types of tumors (as described in Chapter 2); and (3) several cancer gene pathways, including RAS and PI3K/AKT, converge on the C-MYC protein and directly enhance its stability. In each case, an increased concentration of C-MYC protein promotes the transcriptional transactivation of C-MYC target genes.

p53 Activation is Triggered by Damaged or Incompletely Replicated Chromosomes

The role of cancer genes in cellular pathways and higher-order networks is well illustrated by the case of p53. As discussed in several previous chapters, mutations that inactivate *P53* are highly prevalent in many cancers. In accordance with this central role in tumor suppression, p53 has been found to participate in many key functions related to cell proliferation and the maintenance of tissue homeostasis. The p53 protein is activated by diverse upstream signals, and triggers key downstream responses.

Unlike the numerous signal transduction pathways that originate at the cell surface, the p53 pathways signal primarily in response to intracellular stimuli. The ultimate trigger of p53 activation is a change in chromosome structure. Such changes can be the result of either DNA damage or interrupted DNA replication or repair.

Many perturbations in a cellular microenvironment can lead directly or indirectly to the damage of chromosomal DNA. Exposure to ionizing radiation generates highly reactive species that can directly create single and double strand DNA breaks, the latter of which represent a potentially lethal form of DNA damage.

More pernicious types of environmental agents act more indirectly to affect chromosome integrity. A particularly vulnerable period in the division cycle of a cell is S-phase, the interval in which chromatin is replicated. Many metabolic states can cause inhibition of DNA replication and result in the accumulation of DNA replication intermediates. Various agents that damage the DNA template or limit the building blocks of chromatin can cause the stalling of the molecular machines that replicate chromosomes. There are many ways that this occurs. Agents such as ultraviolet light and many types of toxins can create DNA adducts. Such adducts represent physical obstacles that can effectively block replication fork progression. Nutrient deprivation and the accumulation of metabolic byproducts can affect the pathways of biosynthesis that produce chromatin precursors, thereby impeding the replication of the genome. Low levels of oxygen, a state called *hypoxia*, can also inhibit DNA replication and cause increases in p53 levels. In summary, diverse

conditions that make for suboptimal DNA replication can be manifest as DNA breaks or stalled replication forks. These DNA structures can activate the enzyme-mediated pathways that regulate p53.

The kinases that function upstream of p53 are highly sensitive to changes in chromosome structure. While additional types of pathways may affect p53, the most firmly established pathways all function to signal physical alterations of chromosomes. Given its central role in cancer and in DNA damage signaling, p53 has been called the '*guardian of the genome*'.

p53 activity is controlled by coupled phosphorylation and ubiquitination reactions. Under conditions that favor cell growth, the majority of p53 molecules in the cell are physically bound to the protein product of the *MDM2* proto-oncogene (see Fig. 5.25). The MDM2 protein is part of a multienzyme process that adds ubiquitin moieties to proteins. Once modified with ubiquitin, substrate proteins are tagged for degradation by the proteosome. As a result of this degradation cycle, the half-life of p53 in the cell is approximately 5–20 min. The interaction with MDM2 therefore is a mechanism for keeping p53 levels low under conditions in which growth and survival are favored.

Changes in the intracellular or extracellular environment that cause chromosomal breaks or stalling of DNA replication lead to the phosphorylation of p53 by upstream kinases. Phosphorylation of p53 at specific serine residues causes a conformational change that disrupts the p53-MDM2 interaction. The disruption of p53-MDM2 complexes prevents ubiquitin-mediated degradation, causes a several-fold increase in protein half-life and thereby leads to an increase in the intracellular

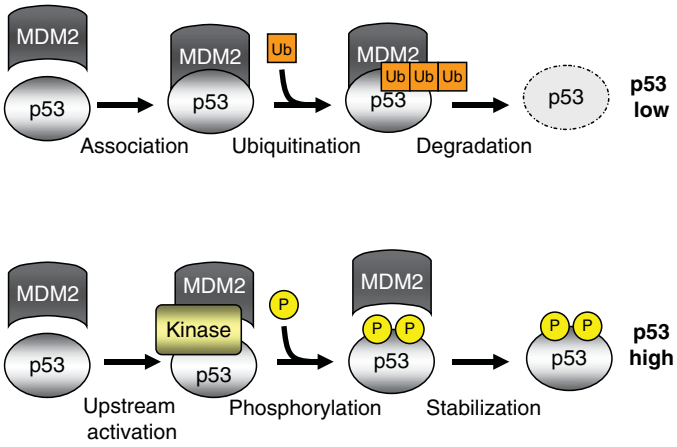

Fig. 5.25 p53 is stabilized upon phosphorylation. Under most conditions, p53 is associated with the MDM2 protein. MDM2 is a component of an enzyme complex that convalently attaches ubiquitin moieties to p53. p53 is thus marked for degradation by the proteosome. By this mechanism, very low levels of p53 are maintained. Under cellular conditions that alter chromosome structure, upstream kinases are activated and catalyze the phosphorylation of p53 at multiple sites. These modifications prevent association of p53 with MDM2, and thus prevent p53 degradation

concentration of p53. The phosphorylation of p53 on multiple sites by upstream kinases directly regulates its activity by controlling its abundance.

The control of p53 concentration by MDM2 is an important mechanism of p53 regulation, but there are clearly additional ways in which covalent modifications affect p53 function. In response to various types of growth inhibitory stress, p53 can be phosphorylated on as many as ten different serine residues and acetylated on at least two lysine residues. In addition, several serine residues have been found to be dephosphorylated by specific phosphatases. Many of these modifications appear to affect p53 conformation and activity in ways that are MDM2-independent. Interestingly, the posttranslational modifications of p53 all appear to occur close the N- and C-termini. The central domain of p53, which is encoded by the exons that are most commonly mutated in cancers, apparently remains largely free of covalent modifications.

What are the kinases that regulate p53, arguably the most important tumor suppressor gene product? From an experimental standpoint, identifying the kinase or kinases that are responsible for the phosphorylation of any given substrate can be a difficult task. Cell-free biochemical systems have been widely used to establish kinase-substrate interactions. However, such relationships that are determined in vitro are in many cases not accurately representative of what actually occurs in vivo. Making matters still more complex is the fact that some protein kinases can modify many diverse substrates both in vitro and in vivo. This characteristic, often referred to as *enzymatic promiscuity*, makes it difficult to determine which substrates are physiologically relevant to the function under investigation. In the case of p53, in vitro studies have identified a number of kinases that can effectively modify p53 on the residues known to be affected by DNA damage in vivo. The roles and relative importance of these kinases remains an important question.

A major breakthrough in understanding the mechanisms of p53 regulation was made in 1992, when Michael Kastan and his colleagues reported that cells from patients affected by ataxia telangiectasia (AT) were defective in their upregulation of p53 after treatment with ionizing radiation. It was subsequently demonstrated that the gene mutated in AT, *ATM*, encodes a kinase that directly phosphorylates p53 on serine 15 in vivo and in vitro. This phosphorylation site is within the p53 domain that physically interacts with MDM2. This series of experiments proved that ATM could phosphorylate p53 and that this single modification has a strong affect on p53 activation. *ATM* and *P53* are tumor suppressor genes. The demonstration of a direct functional interaction between their encoded proteins indicates that *ATM* and *P53* suppress tumorigenesis by participating in a common pathway.

ATM kinase activity is rapidly activated after cells are exposed to DNA-damaging agents that cause double strand DNA breaks, such as ionizing radiation. Detailed biochemical investigation has revealed that ATM is activated at the site of DNA breaks in two defined steps (see Fig. 5.26). When inactive, ATM molecules exist in bound pairs, known as homodimers. Within moments after a DNA break occurs, ATM homodimers dissociate into active monomers. However, these monomers are not fully activated until they interact with a multiprotein complex known as the MRN complex (encoded by *MRE11, NBS1* and *RAD50*), and with the protein encoded by

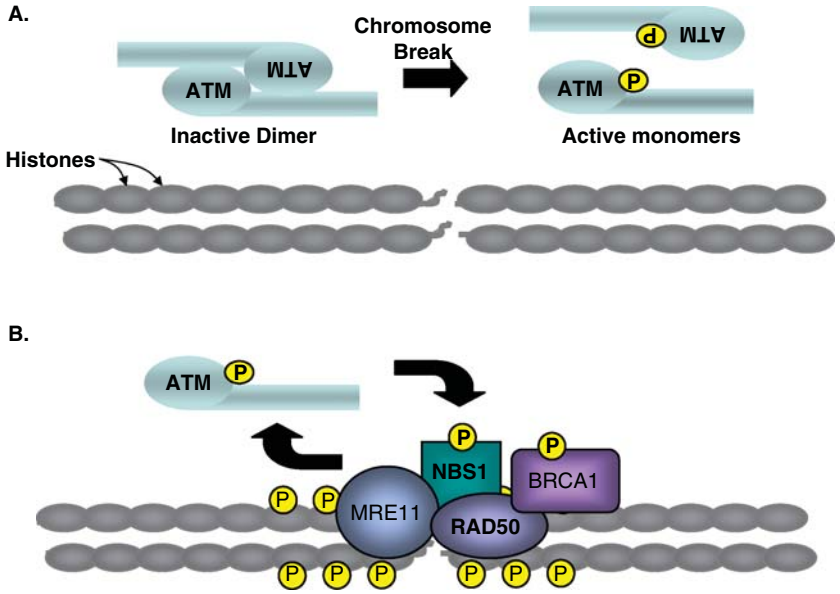

Fig. 5.26 Activation of ATM at the site of a double strand DNA break. Endogenous cellular processes, such as stalled replication forks, and exogenous environmental agents, such as ionizing radiation, can cause double strand DNA breaks. The changes in chromosome structure caused by such lesions activate ATM. **A.** Initially present in inactive, dimeric form, ATM is concurrently phosphorylated (shown in yellow) and dissociated into monomers. **B.** Thus activated, ATM phosphorylates histones in the vicinity of the double strand break, as well as proteins in the MRN (MRE11, RAD50, NBS1) complex and BRCA1. These protein/DNA complexes actively recruit and retain ATM at the double strand break site, facilitating the activation of downstream effectors, including FA proteins and p53

BRCA1. The MRN complex binds DNA and has multiple biochemical properties that include cutting, unwinding and bridging the ends of the damaged double helix. BRCA1 protein is involved in non-homologous DNA end joining. It appears that the MRN complex and the BRCA1 protein are required for the efficient recruitment and possibly the retention of ATM monomers at double strand DNA break sites.

Activated ATM is a serine/threonine protein kinase that can phosphorylate numerous downstream substrates. Cumulatively, ATM and proteins downstream compose the *DNA damage signaling network*, interconnected molecular circuits that integrate many upstream stimuli and numerous several downstream targets, including p53 (see Fig. 5.27). Via this complex network, perturbations to chromosomes can trigger many diverse cellular responses. Recent studies have determined that over 700 proteins are phosphorylated by ATM and/or ATR kinases, at more than 900 sites.

It has not been firmly established which of the DNA damage-response kinases that can phosphorylate p53 is most important in terms of tumor suppression. Notably, there is not a single kinase gene that when mutated recapitulates all of the cancer-related phenotypes seen upon mutation of *P53*. There are several explanations for this: (1) the kinases that compose the DNA damage signaling network

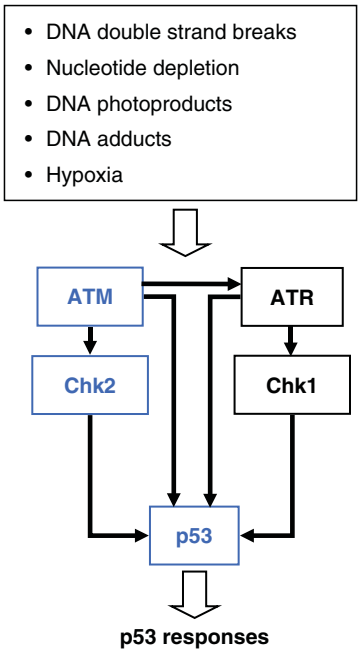

Fig. 5.27 Activation of p53 by the DNA damage response network. Diverse types of environmental agents and cell states can alter chromosomal DNA structure and impair DNA replication. Such stimuli activate ATM and ATR, the apical kinases of the DNA damage response network. Signals generated by ATM and ATR are amplified by the checkpoint kinases, Chk1 and Chk2. All of these kinases can phosphorylate and activate p53 under some conditions. Proteins encoded by established tumor suppressor genes are highlighted in blue

function together to upregulate p53, such that any single one is partially redundant with respect to the others; or (2) different kinases are of primary importance in different tissues, and in response to different types of chromosome-damaging stimuli.

Several of the most highly studied components of the DNA damage signaling network are:

Ataxia telangiectasia- and Rad3-related (*ATR*). *ATR* is a gene in the same family of kinase genes as *ATM*, and encodes a protein kinase that triggers a parallel signaling pathway. *ATR* is not involved in AT, nor is it mutated at a significant frequency in sporadic tumors. In contrast to ATM, which is activated by double strand DNA breaks, ATR is activated by DNA structures that accumulate when DNA replication forks stall. ATR is therefore triggered by ultraviolet light and other types of agents that cause DNA adducts, as well as by agents that deplete the nucleotide pools required for DNA replication. The activity of ATM at double strand breaks can also create DNA structures that are recognized by ATR. For this reason, ATM is upstream of ATR in the responses to stimuli such as ionizing radiation, but in a parallel pathway with respect to stimuli that directly inhibit DNA replication.

Checkpoint kinase 2 (CHK2). CHK2 encodes a serine/threonine kinase that is phosphorylated and thus activated by ATM. Chk2 is believed to functionally amplify the ATM signal. Under some conditions, Chk2 can phosphorylate p53 on serine 20. Germline *Chk2* mutations have been found in up to 1% of some populations, and have been associated with small increases in cancer risk. The current view of Chk2 is that it is a low penetrance tumor suppressor gene.

Checkpoint kinase 1 (CHK1). CHK1 was originally discovered as a homolog of a checkpoint kinase gene in yeast that acts as a central regulator of growth control,

particularly after DNA damage. Structurally unrelated to Chk2, Chk1 is a serine/threonine protein kinase that is a component of the pathway that includes ATR. ATR directly phosphoryates Chk1 on several residues. Thus activated, Chk1 can phosphorylate p53 on sites that are important for its upregulation. Chk1 is not mutated in a significant proportion of tumors in any type of cancer.

P53 Induces the Transcription of Genes that Suppress Cancer Phenotypes

P53 encodes a transcription factor. Unlike the transcription factor encoded by *C-MYC*, p53 primarily regulates genes that inhibit growth and facilitate cellular repair processes.

When stabilized and activated by posttranslational modifications, p53 becomes more abundant and assembles into tetramers. In this configuration, p53 binds tightly to DNA that contains the p53 consensus DNA-binding sequence (see Fig. 5.28). This sequence motif is commonly found in the promoter regions of genes. Thus bound to a promoter element, p53 is a strong inducer of gene expression.

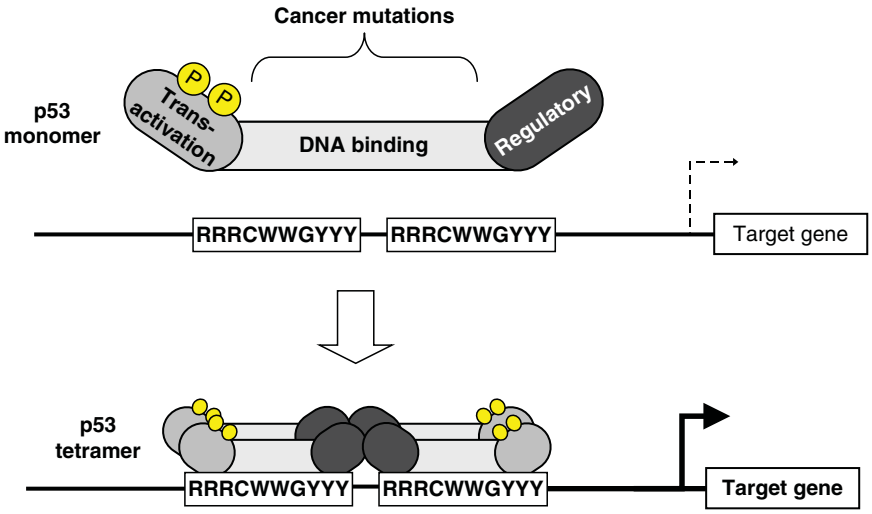

Fig. 5.28 p53 transactivates transcription. The p53 protein contains a central core domain that is required for sequence-specific DNA binding, an N-terminal domain that interacts with other transcription factors and facilitates transactivation, and a C-terminal domain that is required for protein oligomerization and regulation. The DNA-binding domain is encoded by the *P53* regions that are most commonly mutated. The N- and C-termini contain the residues that are modified by upstream kinases and acetylases. Shown are the phosphate groups that disrupt MDM2 binding. When activated, p53 assembles into tetramers. Tetrameric p53 binds to a promoter element containing two half sites that contain the RRRCWWGYYY DNA sequence motif (R = purines A or G; W = A or T; Y = pyrimidines C or T), separated by 0–13 nonspecific bases (only one DNA strand is shown). The binding of activated p53 increases promoter activity (arrow) and target gene expression

Almost all of the single nucleotide substitutions that inactivate *P53* are located in the exons that encode the central, DNA-binding domain. Proteins encoded by cancer-associated mutant *P53* genes therefore fail to bind promoter sequences and also fail to transactivate transcription. While other biochemical functions have been attributed to the p53 protein, it is the transcriptional transactivation function that is most universally inactivated in cancers that harbor *P53* mutations. The preponderance of mutations in the DNA-binding region suggests that the tumor suppressor function of p53 is dependent on its role as a transcription factor.

Many genes can be turned on by activated p53. There are an estimated 1,600 copies of the p53 DNA-binding consensus sequence scattered throughout the genome, although the majority of these do not appear to be in positions that would affect the transcription of protein-coding genes. Nonetheless, many different genes are essentially switched on as a result of the binding of activated p53 to gene promoter sequences. The gene transcripts that are induced by activated p53 can be collectively referred to as the *p53-dependent transcriptome*.

The p53-dependent transcriptome encodes proteins that function in downstream pathways that regulate cell birth, growth and death (see Fig. 5.29). Cancer-associated mutations that inactivate p53 disrupt these pathways and thereby cause significant changes in the ways that mutant cells respond to environmental stressors that affect chromosome structure. The normal cellular response to stressful changes in the microenvironment is to stop growing and to attempt to repair DNA. Cell clones that have lost p53 function continue to grow.

Several target genes and pathways provide a representative view of the functions that are lost when p53 is inactivated during tumorigenesis:

Cell cycle arrest. Several p53 target genes function to control the progression of the cell division cycle. There are several mechanisms by which p53 can exert this

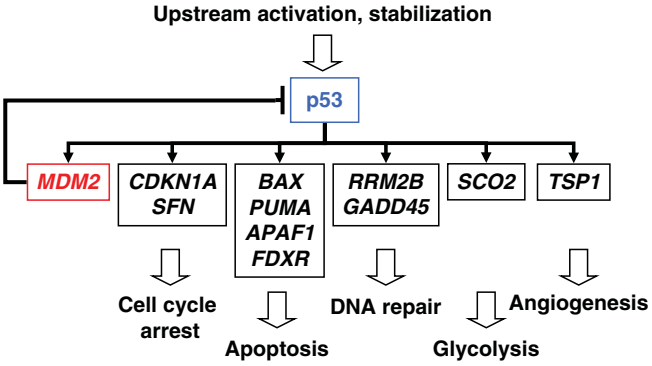

Fig. 5.29 Genes induced, and pathways activated, by p53. Following its stabilization and activation, p53 transactivates the transcription of numerous target genes. These genes, in turn, control many aspects of cell proliferation. By regulating the expression of the proto-oncogene *MDM2*, p53 effectively regulates its own activity via a feedback loop. The genes shown are a representative subset of the genes known to be regulated by p53

control. The p21 protein, encoded by *CDKN1A*, binds and thereby inactivates several of the cyclin-dependent kinases that drive the cell cycle forward. The 14-3-3σ protein, encoded by *SFN*, alters the intracellular localization of cyclin-dependent kinase regulators. Cells with mutant *P53* alleles are deficient in their ability to arrest the progression of the cell cycle.

Programmed cell death (apoptosis). p53 induces several genes that function in the pathways that mediate apoptosis. Proteins such as Bax, ferrodoxin reductase and PUMA function to destabilize mitochondria and thereby lower the threshold of apoptotic stimuli. The APAF1 protein derepresses the apoptotic proteases that are triggered by upstream signals. Overall, loss of p53 function *decreases* the ability of cells to undergo apoptosis.

DNA repair. p53 participates in several processes that are triggered as cells attempt to repair damaged DNA. The *RRM2B* gene encodes a subunit of ribonucleotide reductase, an enzyme that is required for mobilizing the nucleotides required for DNA repair. The protein encoded by *GADD45* promotes nucleotide excision repair and blocks the activity of proteins at the replication fork, inhibiting DNA synthesis. This activity presumably facilitates the coordination of cell cycle arrest and DNA repair.

Glucose metabolism. Normal cells employ the highly efficient process of *aerobic respiration* to convert the energy of glucose into adenosine triphosphate (ATP). In contrast, cancer cells preferentially rely on the anaerobic pathway known as *glycolysis* to provide ATP. This metabolic change is termed the Warburg effect (see Chapter 1). The Warburg effect provides a distinct selective advantage to cells growing in regions of low oxygen concentration, such as those that occur around growing tumors. *SCO2* encodes the synthesis of cytochrome C oxidase 2 protein, a regulator of the cytochrome oxidase C complex, the major site of oxygen utilization in eukaryotic cells. *SCO2* expression is upregulated by p53. The loss of p53 function reduces *SCO2* expression, resulting in a switch in metabolism from respiration to glycolysis. Cells with mutated *P53*, therefore have a survival advantage in regions of low oxygenation.

Angiogenesis. As they increase in size, solid tumors begin to trigger the growth of new blood vessels, which facilitate further growth and metastasis. This critical process is known as *angiogenesis*. Normal tissues secrete factors that inhibit angiogenesis. One of the first such inhibitors to be discovered is a protein called thrombospondin, which is encoded by *TSP1*. Expression of *TSP1* is stimulated by p53. Therefore, loss of p53 function leads to the decreased expression of *TSP1* and the disinhibition of angiogenesis.

The MDM2-p53 Feedback Loop

Cancers that harbor amplified *MDM2* alleles generally do not have *P53* mutations. Why is this the case? The answer lies in the biochemical relationship between p53 and MDM2. The MDM2 gene is a target of p53 transcriptional transactivation. As is the case with other p53 target genes, *MDM2* contains the p53 DNA-binding consensus sequence in its promoter, and can be transcriptionally activated by

bound p53 tetramers. The induction of MDM2 by p53 has significant implications for p53 activity, and additionally explains the pattern of mutations that affect *MDM2* and *P53*.

The activation of *MDM2* expression by p53 constitutes a negative feedback loop that functions to attenuate the p53 response. As described above, the levels of p53 in the cell are largely controlled by the association between p53 and MDM2 protein. Initially, the phosphorylation of p53 by upstream kinases disrupts MDM2 binding and stabilizes p53. Subsequently, the upstream signals subside and MDM2 expression increases due to p53 transactivation. The increased levels of MDM2 protein cause a reduction in p53 levels and a concomitant downregulation of p53 activity (see Fig. 5.30).

The MDM2-p53 feedback loop provides a critical insight into how *MDM2* functions as an oncogene. In some types of cancer, most commonly sarcomas, *MDM2* is converted from a proto-oncogene to an oncogene by gene amplification. As a result of this type of mutation, *MDM2* is expressed at significantly higher levels, even in the absence of p53. The high levels of MDM2 protein in cancer cells with *MDM2* amplification inhibit p53 function. Thus, *MDM2* amplification phenocopies inactivating mutations of *P53*.

Cancer cell clones containing specific mutations expand because those mutations provide a selective advantage. In cells that acquire *MDM2* amplification, p53

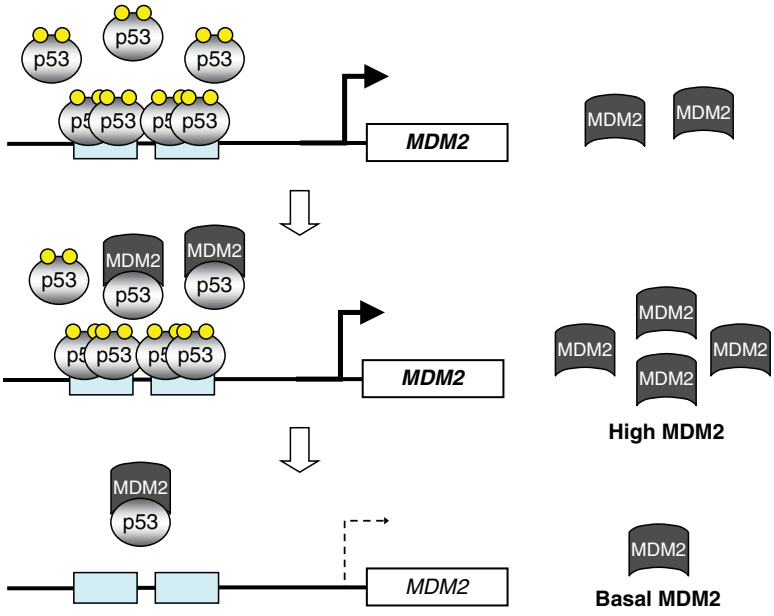

Fig. 5.30 Attenuation of p53 activity by MDM2. Phosphorylated p53 is stabilized and assembles into tetramers that bind the p53 consensus binding sites (blue) in the *MDM2* promoter. High levels of MDM2 protein thus accumulate. As upstream signals subside, MDM2 associates with unbound p53 molecules, targeting them for degradation. Lowered levels of p53 result in the decrease in *MDM2* expression to basal levels

becomes functionally downregulated. There would therefore be little additional advantage to clones that subsequently acquire mutated *P53*.

The mutually exclusive relationship between *MDM2* amplification and inactivating *P53* mutations illustrates a central principle of cancer genetics. Cancer cells invariably have many mutations. However, concurrent mutations within a single pathway are rarely seen because once a mutation disrupts a pathway, there is no additional selective pressure for loss of other genes that function in that pathway. This theme recurs throughout the pathways populated by cancer genes.

The DNA Damage Signaling Network Activates Interconnected Repair Pathways

The genes that cause the rare, autosomal recessive diseases that link genetic instability with cancer (discussed in Chapter 4) are generally involved in the repair of damaged DNA. Different types of DNA damage are repaired by distinct multiprotein complexes. The activity of repair complexes is stimulated by the same DNA damage signaling network that causes the upregulation of p53 (see Fig. 5.31).

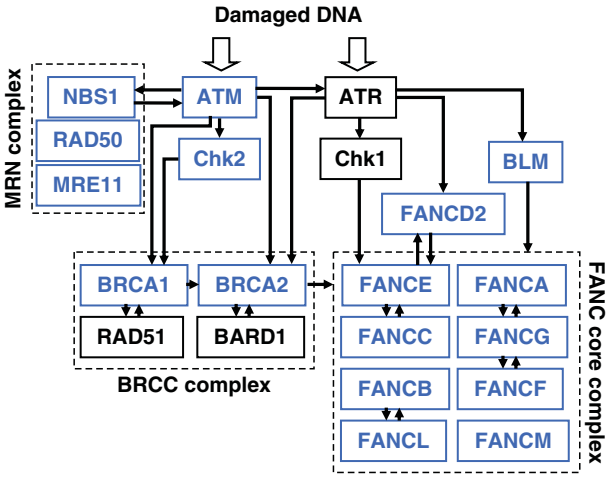

Fig. 5.31 The DNA damage signaling network is connected to DNA repair complexes. The DNA damage responses triggered by activation of ATM and ATR kinases involve multiple pathways that contain cancer genes. Components of the Fanconi anemia (FA) core complex, the Mre11/Rad50/NBS1 (MRN) core complex, and the BRCA1/BRCA2 containing (BRCC) complex are directly activated by upstream signaling kinases. Insight into the physical relationships between the components of these complexes has been gained by biochemical studies that have revealed multiple pairwise associations (as shown). Many additional proteins and protein–protein interactions have been reported, only the most highly illustrative of these are shown. Proteins highlighted in blue are encoded by genes that, when mutated, are associated with cancer predisposition

The cellular hallmark of Fanconi anemia (FA) is a marked sensitivity to the effects of DNA crosslinking agents. Accordingly, the FA gene-encoded proteins have been found to function in repair complexes that process DNA interstrand crosslinks. Of the 11 FA genes that have been identified, eight are components of a multiprotein FA core complex. The precise role of each of the FA proteins remains unclear, but several contain protein motifs that suggest distinct biochemical functions. These include FANCJ, which contains a helicase domain and FANCL, which contains ubiquitin ligase activity. The FA core complex, formed by FANCA, FANCB, FANCC, FANCE, FANCF, FANCG, FANCL and FANCM, functions in the nucleus to add a single ubiquitin moiety to the FANCD2 protein. The monoubiquitination of FANCD2 allows it to associate with chromatin at sites of DNA damage and repair. At these sites, FANCD2 colocalizes with ATR, and the helicase encoded by *BLM*. The exact role of FANCD2 at the chromatin repair site is unclear.

Also located at repair foci, and thus implicated in the DNA repair process, are BRCA2, BRCA1 and NBS1. BRCA2 is partnered with RAD51, a protein required for homologous recombination-mediated DNA repair. BRCA1 associates with BARD1, which appears to play a role in the monoubitquitation of FANCD2. The BRCA1 and BRCA2 genes function in a distinct multiprotein complex that has been termed the BRCA1- and BRCA2-containing (BRCC) complex. NBS1, along with MRE11 and RAD50 is a component of the MRN complex. This protein complex is important for the recruitment and retention of ATM to the sites of DNA double strand breaks.

Many genes involved in DNA repair and metabolism are mutated in recessively-inherited cancer predisposition syndromes. These findings reveal the central importance of DNA repair to the maintenance of genetic stability and the suppression of tumors. However, whether these genes are at all involved in the common forms of cancer is unclear. The recessive mutations that cause DNA repair/DNA instability syndromes, e.g. XP, AT, BS, FA, and WS, are not found in a significant number of sporadic tumors. It is possible that subtle and varied alterations in DNA repair genes may contribute to lesser degrees of cancer predisposition that have yet to be detected by population-based studies.

Inactivation of the Pathways to Apoptosis in Cancer

The maintenance of tissue homeostasis depends not only on the rate of cell proliferation, but also on the rate of cell death. The stability of adult tissues is largely dependent on highly conserved signaling pathways that cause a form of cell death known as *apoptosis*. In contrast with *necrosis* that results from physical or chemical insult and which triggers inflammatory responses, apoptosis is genetically programmed and occurs in the absence of any apparent physical trauma.

The selective elimination of cells by apoptosis is critical to processes as diverse as development and the modulation of immune responses. Apoptotic pathways

appear to be functional in all normal cells of the human body. In many cancers, the pathways that lead to apoptosis are disrupted by either proto-oncogene activation or by tumor suppressor gene inactivation. As a result, cancer cells fail to respond normally to death signals and may also be relatively resistant to the effects of anti-cancer therapy.

Two distinct categories of stimuli can cause apoptosis. Extracellular signaling molecules, including highly specific cytokines, hormones and growth factors, can activate what is known as the *extrinsic pathway*. Via this pathway, cell surface receptors generate signals upon binding death-inducing ligands (see Fig. 5.32). The receptors that can trigger apoptosis include the tumor necrosis factor (TNF) receptor superfamily, the TNF-related apoptosis inducing ligand (TRAIL) and Fas. The use of recombinant ligands to activate these receptors and induce apoptosis in cancer cells will be described in Chapter 7.

Alternatively, death receptor-independent apoptosis can be caused by various types of cellular stress, including chromosomal DNA damage and failure to complete mitosis. Such intracellular events activate what is known as the *intrinsic pathway* to apoptosis (see Fig. 5.33). A key component of the intrinsic pathway is the mitochondrion. Indeed, the intrinsic pathway is sometimes referred to as the mitochondrial pathway.

The main site of cellular ATP generation, the mitochondrion is also the target of both anti- and pro-apoptotic mediators which respectively function to stabilize and destabilize the mitochondrial outer membrane. Activation of the intrinsic pathway causes disruption of the mitochondrion and the release of molecules present in the intermembrane space into the cytoplasm. Among the molecules normally sequestered

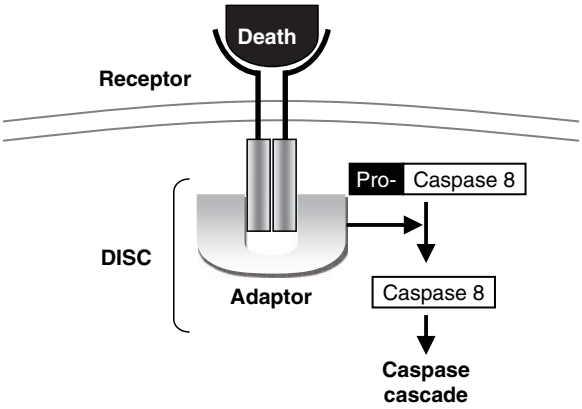

Fig. 5.32 The extrinsic pathway to apoptosis. Also known as the death receptor pathway, the extrinsic pathway is triggered by extracellular ligands. Receptor molecules – that include the TNF-receptor superfamily, Fas, and the TNF-related apoptosis inducing ligand (TRAIL) – multimerize in response to death ligand binding. Adaptor proteins associate with the intracellular domains of complexed receptors and recruit the pro-enzyme form of caspase 8. Together, these components form the death-inducing signaling complex (DISC) that triggers the downstream cascade of caspase activation

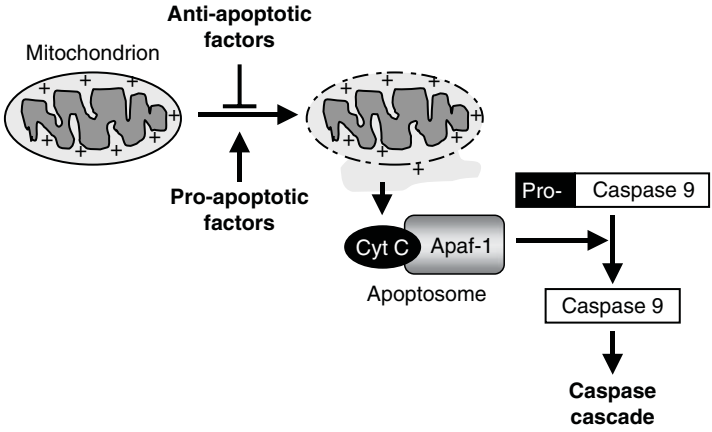

Fig. 5.33 The intrinsic pathway to apoptosis. Apoptotic stimuli cause the destabilization of the mitochondrial outer membrane. Disruption of the mitochondrion results in the release into the cytoplasm of the contents of the mitochondrial intermembrane space, including charged radicals (+) and cytochrome C. Cytochrome C forms a complex with an apoptosis activating factor protein called Apaf-1. This structure, known as the apoptosome, facilitates the cleavage and activation of caspase 9. Mitochondrial stability is controlled by both anti- and pro-apoptotic factors

in the mitochondrion are reactive oxygen species and the electron transport protein cytochrome C. When released into the cytoplasm, cytochrome C binds to a protein called Apaf-1, forming a complex known as the *apoptosome*.

The outcome of activation of either of the two apoptosis pathways is the onset of marked cellular changes that include disruption of cellular and nuclear membranes and the breakdown of chromatin. Cellular proteins are digested by the *caspases*: cysteine proteases that cleave polypeptides at aspartic acid residues. Many of the characteristic features of apoptosis can be blocked by chemical inhibitors of the caspases, demonstrating their central role as the major effectors of apoptotic pathways.

Caspases are translated as pro-enzymes called *pro-caspases*. Cleavage of pro-caspases results in their activation. The DISC (see Fig. 5.32) facilitates the proteolytic activation of caspase 8 by the extrinsic pathway, while the apoptosome (see Fig. 5.33) causes the proteolytic activation of caspase 9 by the intrinsic pathway. These upstream caspases, also known as *initiator caspases*, then cleave and activate downstream caspases, also known as *effector caspases*. The rapid and irreversible activation of the effector caspases by the initiator caspases, a process known as the *caspase cascade*, ultimately results in the proteolytic degradation of the cell.

The pathways to apoptosis are frequently inhibited by mutations that drive tumorigenesis (see Fig. 5.34). Most commonly, cancer genes affect the intrinsic pathway and thus disable the apoptotic responses that are normally triggered when cells are damaged or fail to divide properly.

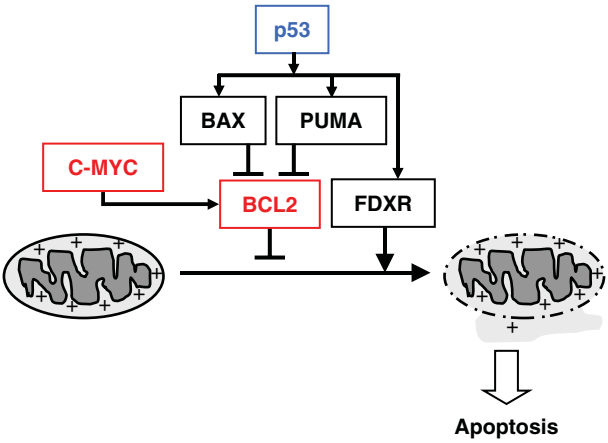

Fig. 5.34 Cancer genes affect mitochondrial stability. Oncogenic pathways that induce C-MYC dependent transcription cause an increase in levels of the apoptosis inhibitor BCL2. BCL2 is inhibited by the induction of the proapoptotic BCL2-family members BAX and PUMA, which are induced by p53. p53 also increases the expression of ferredoxin reductase (FDXR), which directly destabilizes the mitochondrial outer membrane

The regulator of mitochondrial membrane stability first identified is the protein encoded by the *BCL2* gene. *BCL2* was originally cloned in 1984, by Carlo Croce and coworkers, as an oncogene activated by a common translocation in B-cell lymphomas. While all oncogenes known at that time could be demonstrated to cause increased cell proliferation when experimentally introduced into cultured cells, *BCL2* did not have this expected effect. Upon further study, overexpression of *BCL2* was shown to confer resistance to stimuli that would otherwise cause apoptosis. Thus, *BCL2* defined a new type of oncogene, one that functioned not by increasing proliferation, but by inhibiting cell death.

BCL2 is a member of an eponymous family of highly interactive proteins that control apoptosis by affecting the stability of the mitochondrial membrane. The BCL2 family contains both pro-apoptotic and anti-apoptotic members, all of which share characteristic protein motifs. It is the balance between the two types of BCL2 proteins that determines the threshold for mitochondrial destabilization. While BCL2 proteins have been detected in close physical proximity to the outer mitochondrial membrane, the exact mechanism by which these proteins affect mitochondrial stability is unclear. Interestingly, structural homologies have been observed between BCL2 family proteins and bacterial proteins that function in membrane pore formation.

In addition to its many other functions, p53 plays a major role in apoptosis. Apoptotic signals cause the stabilization and activation of p53, which then functions as a potent mediator of mitochondrial stabilization. Among the apoptotic targets of activated p53 are the BCL2 family members BAX and PUMA. These pro-apoptotic

proteins antagonize the anti-apoptotic effects of BCL2. p53 also transactivates transcription of the *FDXR* gene. *FDXR* encodes ferredoxin reductase, which controls mitochondrial membrane stability independently of the BCL2 family.

RB and the Regulation of the Cell Cycle

The counterbalance to cell death, in terms of tissue homeostasis, is cell proliferation. Proliferating cells undergo repeated iterations of growth and division, a process known as the *cell cycle*. The cell cycle is composed of four discrete phases (see Fig. 5.35). The replication of the genome occurs during S-phase, a period of DNA synthesis. Chromosomes are segregated and cells physically divide into daughter cells during mitosis. Cells increase in mass during two gap phases called G_1 and G_2 that occur prior to S-phase and prior to mitosis, respectively. It is during G_1 and G_2 phases that cells are highly responsive to proliferative and antiproliferative stimuli.

The basic mechanisms by which cells progress from one phase of the cell cycle to subsequent phases have been elucidated in model organisms, including yeasts, amphibians and sea urchins. Pioneering studies conducted during the 1970s and 1980s revealed that cells in different phases of the cell cycle have characteristic patterns of protein kinase activity. Phase-specific kinase activation is initiated by a class of proteins called *cyclins*, so named because the prototypes were found to increase and decrease in abundance in cyclical fashion coinciding with the phases of the cell cycle. During each of the cell cycle phases, a characteristic cyclin binds and activates a distinct serine/threonine protein kinase called a *cyclin-dependent kinase* (CDK). Thus activated, the cyclin/CDK complex phosphorylates phase-specific substrates (see Fig. 5.36). For example, G_1 cyclin/CDKs prepare the cell to undergo DNA replication; S-phase-specific cyclin/CDKs promote progression of replication forks and coordinate the firing of replication origins; mitotic cyclin/CDKs promote the creation of the mitotic spindle and the concomitant dissolution of the nuclear membrane. These mechanisms of cell cycle progression are highly conserved in human cells.

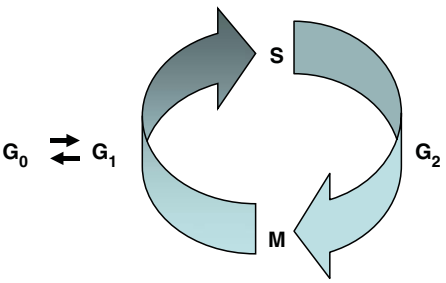

Fig. 5.35 The phases of the cell cycle. DNA replication occurs during S-phase (S). Mitosis (M) results in chromosome segregation and cell division. These two phases are separated by two gap periods (G_1 and G_2), during which increases in cell mass occur. Cells can exit the cell cycle from G_1 and enter a nonproliferative state known as G_0

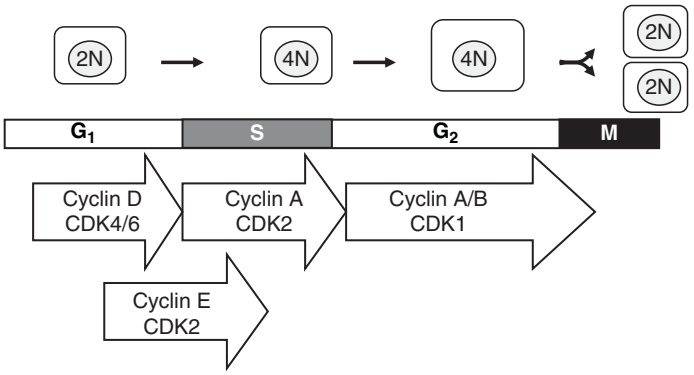

Fig. 5.36 Cell cycle progression is driven by sequentially activated cyclin-dependent kinase complexes. A single cell cycle is illustrated in linear form. Cells in G_1 have a diploid DNA content (2N) which is replicated during S phase. Transition between phases of the cell cycle is driven by the sequential assembly and activation of cyclin/CDK complexes. Each activated cyclin/CDK complex phosphorylates phase-specific substrates. Note that several cyclins and CDKs can associate with different partners at during successive phases. Shown are the most highly characterized CDKs; at least eight CDKs have been identified

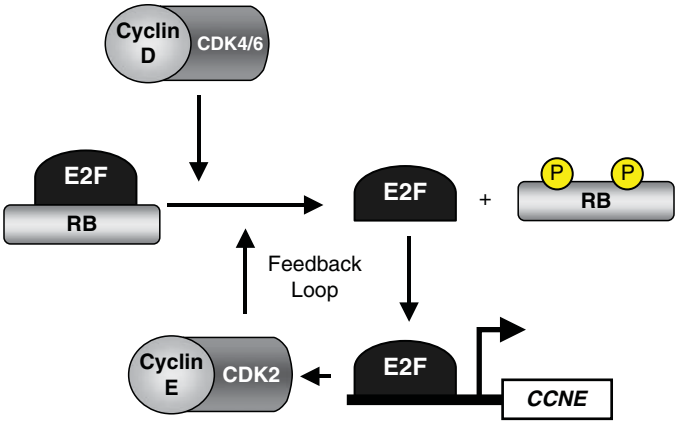

Fig. 5.37 Activation of cyclin E expression by cyclin D-dependent RB phosphorylation. During G_1, RB is bound to the transcription factor E2F. Phosphorylation of RB by cyclin D/CDK4/6 complexes disrupts this association and thus frees E2F to stimulate the transcription of cyclin E. The accumulating cyclin E/CDK2 complexes also phosphorylate RB, forming a positive feedback loop that results in the progression of the cell into S phase

The first molecular connection between cancer genes and the regulation of the cell cycle was made with the cloning and characterization of *RB*, the gene mutated in retinoblastoma (Chapter 3). The RB protein plays a critical role in the regulation of the $G_1 \rightarrow S$ transition (see Fig. 5.37). RB directly controls the activity of E2F,

which promotes the expression of several genes required for S-phase, including the gene that encodes cyclin E, the primary S-phase cyclin. Prior to S-phase, E2F is bound to unphosphorylated RB protein. In this bound form, E2F does not associate with gene promoters; it is functionally sequestered by RB.

Transcription of the cyclin D gene *CCND1* is highest in early G_1 phase. By late G_1, accumulating cyclin D stimulates the activation of CDKs 4 and 6, which then phosphorylate RB. Phosphorylation of RB disrupts the RB-E2F complex, thus allowing E2F to transactivate its target genes, including *CCNE* (the gene that encodes cyclin E). Cyclin E/CDK2 complexes also phosphorylate RB, accelerating the production of cyclin E and the resultant transition into S phase. By this mechanism, the accumulation of cyclin D during G_1 stimulates a subsequent wave of cyclin E expression that drives the cell into S-phase.

Cyclin D proteins are important targets of several mitogenic signaling pathways. In cancers, loss of RB function renders the $G_1 \rightarrow S$ transition cyclin D independent (and therefore mitogen independent). Gain-of-function mutations in genes that encode cyclin D proteins have an equivalent effect. *CCND1* and *CCND3*, which encode two cyclin D family members, are proto-oncogenes that are amplified at low frequency in several types of cancer. *CCND3* is frequently activated by translocation in B-cell lymphomas.

Cancer cells frequently contain increased levels of cyclin D. This observation can be attributed to the regulation of cyclin D levels by several upstream cancer gene pathways. The RAS pathway promotes *CCND1* expression, while the WNT/APC pathway has been shown to inhibit *CCND1* expression via the inhibition of β-catenin. *CCND1* is also a direct target of C-MYC-dependent transcriptional transactivation. The PI3K/AKT pathway prevents the inhibition of targeted cyclin D protein inactivation by the GSK3 kinase. In summary, cyclin D levels can increase via multiple cancer gene pathways, as a result of either the inactivation of tumor suppressor genes or the activation of oncogenes.

Uncontrolled proliferation is one of the universal hallmarks of cancer. The seminal discovery and characterization of RB provided unparalleled insight into how human cells proliferate, and how this process can be dysregulated at the most fundamental level in cancer cells.

Several Cancer Gene Pathways Converge on Cell Cycle Regulators

Cell growth is tightly controlled. Accordingly, the cyclin/CDK complexes that mediate cell cycle transitions are subject to several modes of regulation (see Fig. 5.38). The most basic form of regulation is by the abundance of cyclin. Indeed, CDKs are defined as such by their requirement for cyclin binding.

A second mode of CDK regulation involves postranslational modification. Cyclin/CDK complexes can be inhibited by the phosphorylation of the CDK subunit by highly conserved tyrosine kinases that regulate cell size, including

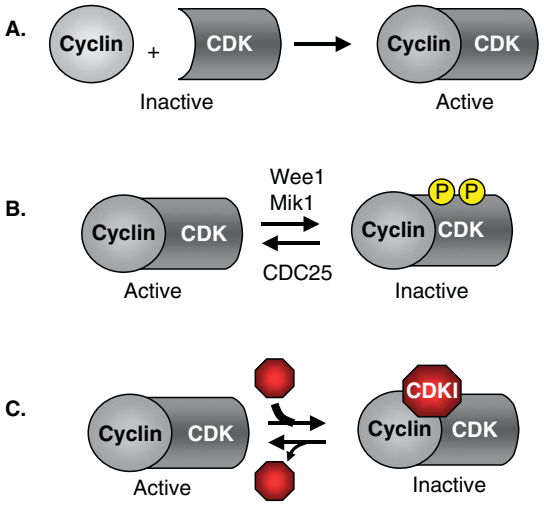

Fig. 5.38 Mechanisms of CDK regulation. **A.** Monomeric cyclin-dependent kinases (CDKs) are essentially devoid of enzymatic activity. The association of cyclins with partner CDKs triggers the transition between cell cycle phases during unperturbed growth. **B.** Cyclin/CDK complexes can be inactivated by tyrosine phosphorylation (shown in yellow). Kinases that limit cell growth, such as wee1 and mkk1, directly phosphorylate CDKs, rendering them catalytically inactive. This inhibition can be relieved by removal of the phosphates by protein tyrosine phosphatases belonging to the CDC25 family. **C.** Cyclin/CDK complexes can be reversibly inactivated by physical association with cyclin-dependent kinase inhibitor proteins, including p15, p15 and p21

Wee1 and Mik1. Phosphorylation on specific tyrosine residues renders CDKs catalytically inactive. These inhibitory phosphates can be removed by the dual-specificity phosphatases belonging to the CDC25 family. The balance between tyrosine kinase and phosphatase activities sets the threshold for CDK activation and cell cycle progression. This mode of regulation appears to be largely intact in cancer cells. Neither the Wee1/Mik1 tyrosine kinases, nor the CDC25 family of phosphatases appear to be differently regulated in cells with commonly mutated cancer genes.

A third means of regulating cyclin/CDK complexes is the binding of inhibitory subunits known as *cyclin-dependent kinase inhibitors* (CDKIs). Human cells express two distinct classes of CDKIs. *Universal* CDKIs associate with all cyclin/CDK complexes and therefore function to inhibit all cell cycle transitions. The three members of this class are designated by molecular weight: p21, p27 and p57. Proteins belonging to the second class of CDKIs bind exclusively to the CDK4 and CDK6 complexes that specifically mediate the transition from G_1 to S phase. These include the two protein products of the *CDKN2A* locus, p16 and p14(ARF), and that of the neighboring gene *CDKN2B*, which encodes p15 (see Chapter 3). These three CDKIs are sometimes referred to as *INK4 proteins*, reflecting their ability to inhibit CDK4.

Many types of cancers develop defects in CDKI pathways. There are two ways in which CDKI dysfunction can arise during tumorigenesis. (1) Mutations can directly disrupt CDKI genes. For example, the two transcripts from *CDKN2A* are frequent targets of mutations in some types of cancer. Cancer cells that harbor inactivating *CDKN2A* mutations will exhibit loss of p14(ARF) and/or p16 function and, thus a loss of control over the $G_1 \rightarrow S$ transition. (2) CDKI pathways can be disrupted by upstream mutations in one of several cancer gene pathways that induce the transcription of CDKI genes. The TGF-β and p53 pathways are two examples of cancer gene pathways that directly control CDKI gene expression (see Fig. 5.39).

Among the target genes of the TGF-β pathway is *CDKN2B*. In most normal cells, TGF-β ligand induces the expression of *CDKN2B* and a concomitant increase in p15 protein. p15 then associates with and inhibits CDK4 and CDK6 complexes, blocking the cell cycle by preventing the $G_1 \rightarrow S$ transition. Cancer cells that have developed mutations that disrupt the TGF-β signaling pathway fail to upregulate p15 and thus have reduced control over entry into S.

A direct connection between p53 and the regulation of the cell cycle became apparent with the discovery, in 1993, that *CDKN1A* is a target of p53 transcriptional transactivation. *CDKN1A* encodes p21, a universal CDKI that regulates multiple cell cycle transitions. In most proliferating cells, p21 protein is present at very low levels. Upon activation by DNA strand breaks or DNA replication intermediates, p53 associates with a binding motif in the *CDKN1A* promoter and dramatically increases *CDKN1A* transcription. Cancer cells that have acquired *P53* mutations during tumorigenesis have impaired induction of *CDKN1A* transcription, and therefore fail to restrict the progression of the cell cycle in response to damaged and incompletely replicated chromosomes.

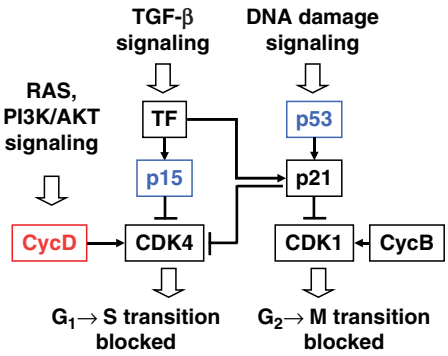

Fig. 5.39 Upregulation of CDKIs by cancer gene pathways. Several cancer gene pathways, including the TGF-β/SMAD and the p53 pathways, induce the expression of CDKI genes. TGF-β ligand results in the upregulation of *CDKN2B* transcription and p15 expression. p15 is an INK4 protein that inhibits the activity of CDK4 and CDK6 complexes, thus blocking transit between G_1 and S phase. p53 activation results in the transcriptional upregulation of *CDKN1A* and increased expression of p21. p21 is a universal inhibitor of CDKs. In contrast, oncogenic pathways such as RAS and PI3K/AKT, induce cyclins that activate CDKs, and thereby promote progression of the cell cycle

Many Cancer Cells are Cell Cycle Checkpoint-Deficient

Genetically programmed growth arrest occurs at defined points in the cell cycle, known as *checkpoints*. First described in yeasts, checkpoints function to ensure that the events of the cell cycle occur in their proper sequence. A checkpoint in G_1 prevents damaged chromosomes from being replicated. A checkpoint in G_2 prevents incompletely replicated chromosomes from being segregated during mitosis. In normal cells, the G_1/S and G_2/M checkpoints provide a means for cells to halt their growth in a coordinated fashion and initiate various modes of DNA repair. In organisms as diverse as humans and yeast, checkpoints function to protect cells from the deleterious effects of failed DNA replication and incomplete chromosomal segregation.

Checkpoints are essentially pathways that functionally inhibit CDKs. As described above (see Fig. 5.38), two modes of CDK inhibition are inhibitory tyrosine phosphorylation and CDKI complex formation. Both of these events are controlled by checkpoint pathways (see Fig. 5.40).

The DNA damage signaling network directly implements checkpoints by the inactivation of CDC25 family members. The most prominent of these is CDC25A. Phosphorylation of CDC25A by upstream kinases results in its rapid degradation

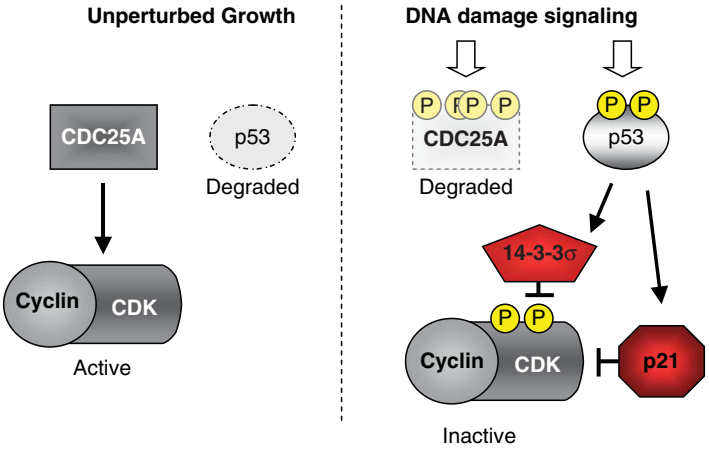

Fig. 5.40 Activation of checkpoints in response to DNA damage. During unperturbed cell growth (left), the CDC25A phosphatase removes inhibitory phosphates from CDKs. Under these conditions, p53 is present at very low levels due to MDM2-dependent degradation. Upon sensing damaged chromosomes, the DNA damage signaling network phosphorylates CDC25A on at least four serine/threonine residues (right). These modifications effectively target CDC25A for degradation by the proteosome. In the absence of CDC25A, the phosphorylated (inactive) form of CDKs quickly becomes predominant. Stabilized, activated p53 induces the CDKI p21 and 14–3–3σ, which sequesters CDKs and prevents them from functioning in the nucleus

by ubiquitin-dependent proteolysis. The loss of a CDK phosphatase tips the balance in favor of CDK inhibitory phosphoryation.

The G_2/M checkpoint is subject to several modes of regulation. The degradation of CDC25A is the first phase of checkpoint activation. The second phase of checkpoint activation is mediated by p21, which accumulates as a result of p53 activation. p21 binds CDK proteins and ensures that they remain inactive, thus stabilizing checkpoint-mediated growth arrest. An additional p53-induced gene, *SFN*, encodes a protein, 14-3-3σ, that functionally sequesters phosphorylated CDK complexes and prevents them from entering the nucleus. That the major checkpoints of the cell cycle are implemented by multiple, overlapping mechanisms is probably an indicator of their importance in maintaining cell viability.

Several cancer gene pathways, including TGF-β/SMAD and the p53 pathway, regulate the cell cycle by activating checkpoints. Some aspects of checkpoint function, namely the induction of CDKI genes, are controlled by cancer gene pathways. The regulation of CDC25 proteins, in contrast, appears to be unaffected by common cancer gene mutations.

During the evolution of cancers, mutations that cause checkpoint deficiency probably allow clonal populations of cells to escape growth arrest normally triggered by signaling molecules or adverse environmental conditions. Thus, checkpoint defects can provide a selective advantage. However, this advantage comes at a price. While checkpoint-deficient cancer cells can escape growth controlling stimuli, they are apparently diminished in their ability to survive more severe forms of DNA damage. Effective DNA repair requires a coordinated halt of cell cycle processes, a function lost in many cancer cells. Indeed, cancer cells have been observed to continue replicating their genomes and to undergo failed mitoses following treatment with DNA-damaging agents such as ionizing radiation. A loss of checkpoint control is thought to underlie the inherent sensitivity of many kinds of cancer cells to the effects of DNA-damaging forms of therapy. The roles of cancer genes in therapeutic responses will be discussed in detail in Chapter 7.

Overview: Dysregulation of Cancer Gene Pathways Confers Selective Advantages

The elucidation of cancer gene pathways has provided insight into how mutated genes cause cancer. Cellular and biochemical studies have revealed how tumor cells are physiologically altered, and how cancer genes that contribute to distinct cancers are functionally interconnected. Studies of cancer genes and the proteins they encode have produced information that is both theoretically enlightening and practically useful.

Mutated tumor suppressor genes and oncogenes serve to highlight the pathways that are particularly important in cancer. The p53 and TGF-β/SMAD pathways, both of which inhibit growth in most cell types, are predominately populated by tumor suppressor genes. Inactivation of one of several tumor suppressors can

therefore lead to disinhibited cell proliferation. Conversely, the RAS and PI3K pathways, which normally function to promote cell growth, are populated by several frequently mutated proto-oncogenes. Oncogenes render growth-promoting pathways constitutively active. Several pathways are populated by both types of cancer genes. In such pathways, mutations that inactivate tumor suppressor genes and mutations that activate proto-oncogenes have opposing effects on pathway function.

In several illustrative cases, cancer gene pathways have provided a road map for the discovery of novel cancer genes. For example, the RAS family of proto-oncogenes was discovered with the use of in vitro transformation assays. Subsequent studies revealed the interaction between RAS and RAF proteins and prompted the close analysis of *BRAF*, which is mutated in cancers at a high frequency. In a similar manner, *PTEN* pointed the way to *PIK3CA*.

As described throughout this chapter, cancer genes populate cellular pathways and complex signaling networks that control cell proliferation and cell death. Traits technically more difficult to quantify, such as cell–cell adhesion, motility and metabolism, are also effected by of cancer gene pathways. Why are these pathways disrupted by mutations in cancers? The answer is that an increased tendency to proliferate and a decreased tendency to undergo cell death provide a selective advantage to the cells that acquire such these traits during tumorigenesis. The acquisition of selectable characteristics defines which cells will become the progenitors of successive waves of expanding cell clones.

With the exception of cells with mismatch repair instability (MMR), neoplastic cells spontaneously acquire mutations throughout the genome in a largely unbiased manner. Cancer genes are not the targets of mutations because they are inherently more mutable than other genes. Rather, cancer genes are propagated by cells that have a proliferative advantage and therefore clonally expand. Ultimately, this occurs because cancer genes change the fundamental characteristics of cells that harbor them.

Further Reading

Bakkenist, C. J. & Kastan, M. B. Initiating cellular stress responses. *Cell* **118**, 9–17 (2004).
Bensaad, K. & Vousden, K. H. P53: New roles in metabolism. *Trends Cell Biol*. **17**, 286–291 (2007).
Blume-Jensen, P. & Hunter, T. Oncogenic kinase signalling. *Nature* **411**, 355–365 (2001).
Cohen, P. The origins of protein phosphorylation. *Nat. Cell Biol*. **4**, E127–E130 (2002).
Dang, V. V. et al. The c-Myc target gene network. *Semin Cancer Biol*. **16**, 253–264 (2006).
Giaccia, A. J. & Kastan, M. B. The complexity of p53 modulation: Emerging patterns from divergent signals. *Genes Dev*. **12**, 2973–2983 (1998).
Giacinti, C. & Giordano, A. RB and cell cycle progression. *Oncogene* **25**, 5220–5227 (2006).
Hajra, K. M. & Fearon, E. R. Cadherin and catenin alterations in human cancer. *Genes Chromosomes Cancer* **34**, 255–268 (2002).
Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).
Horn, H. F. & Vousden, K. H. Coping with stress: Multiple ways to activate p53. *Oncogene* **26**, 1306–1316 (2007).

Kastan, M. B. & Bartek, J. Cell-cycle checkpoints and cancer. *Nature* **432**, 316–323 (2004).
Kastan, M. B. & Lim, D. S. The many substrates and functions of ATM. *Nat. Rev. Mol. Cell Biol.* **1**, 179–86 (2000).
Linding, R. et al. Systematic discovery of in vivo phosphorylation networks. *Cell* **129**, 1415–1426 (2007).
Massague, J. & Gomis, R. R. The logic of TGFbeta signaling. *FEBS Lett*. **580**, 2811–2820 (2006).
Nelson, W. J. & Nusse, R. Convergence of Wnt, beta-catenin, and cadherin pathways. *Science* **303**, 1483–1487 (2004).
Polakis, P. The many ways of Wnt in cancer. *Curr. Opin. Genet. Dev*. **17**, 45–51 (2007).
Samuels, Y. & Ericson, K. Oncogenic PI3K and its role in cancer. *Curr. Opin. Oncol*. **18**, 77–82 (2006).
Sansal, I. & Sellers, W. R. The biology and clinical relevance of the PTEN tumor suppressor pathway. *J. Clin. Oncol*. **22**, 2954–2963 (2004).
Scott, J. D. & Pawson, T. Cell communication: The inside story. *Sci. Am*. **282**, 72–79 (2000).
Sears, R. C. The life cycle of C-myc: From synthesis to degradation. *Cell. Cycle* **3**, 1133–1137 (2004).
Shields, J. M., Pruitt, K., McFall, A., Shaub, A. & Der, C. J. Understanding Ras: 'It ain't over 'til it's over'. *Trends Cell Biol.* **10**, 147–154 (2000).
Simpson, L. & Parsons, R. PTEN: Life as a tumor suppressor. *Exp. Cell Res*. **264**, 29–41 (2001).
Toledo, F. & Wahl, G. M. Regulating the p53 pathway: In vitro hypotheses, in vivo veritas. *Nat. Rev. Cancer* **6**, 909–923 (2006).
Venkitaraman, A. R. Medicine: Aborting the birth of cancer. *Nature* **434**, 829–830 (2005).
Vivanco, I. & Sawyers, C. L. The phosphatidylinositol 3-Kinase AKT pathway in human cancer. *Nat. Rev. Cancer* **2**, 489–501 (2002).
Vogelstein, B. & Kinzler, K. W. Cancer genes and the pathways they control. *Nat. Med*. **10**, 789–799 (2004).
Vogelstein, B., Lane, D. & Levine, A. J. Surfing the p53 network. *Nature* **408**, 307–310 (2000).
Vousden, K. H. & Lane, D. P. P53 in Health and Disease. *Nat. Rev. Mol. Cell Biol*. **8**, 275–283 (2007).
Zhao, J. J. & Roberts, T. M. PI3 kinases in cancer: From oncogene artifact to leading cancer target. *Sci. STKE* **2006**, pe52 (2006).
Zhou, B. B. & Elledge, S. J. The DNA damage response: Putting checkpoints in perspective. *Nature* **408**, 433–439 (2000).

Chapter 6
Genetic Alternations in Common Cancers

Cancer Genes Cause Diverse Diseases

Each of the roughly 100 types of human cancer is caused by the activation of proto-oncogenes and the loss of tumor suppressor genes. Although cancer genomes are complex, some clear mutational patterns are apparent. Several cancer genes are observed very frequently in some types of cancer, but rarely found in other types. Other cancer genes are much more widespread. Recent analysis of individual cancer genomes has shown that, in addition to the well-known cancer genes described in the preceding chapters, there are many mutations that arise, via clonal selection, at very low frequency during tumorigenesis. These observations imply that there are many potential combinations of cancer genes that can cooperatively allow the growth of neoplasia.

The relative importance of a cancer gene or pathway in a given cancer type can be gauged by the frequency at which it is mutated or dysregulated, respectively. For example, mutations in *APC* are found at very high frequency in colorectal cancers, but are only rarely seen in other malignancies. The selection of characteristic mutations of *APC* or other components of the WNT/APC pathway in colorectal cancer probably results from an absolute barrier to neoplastic growth that is highly tissue-specific. In contrast, common mutations affecting *P53* and *K-RAS* have been observed at significant frequency in many types of cancer. From these observations, one can infer that loss of *p53* and loss of control over RAS signaling each provide a selective advantage in many tissue types.

Studies of diseases such as retinoblastoma and colorectal cancer have provided fundamental insights into the nature of cancer genes. Each of these cancers has unique attributes that have facilitated genetic analysis. Retinoblastoma is a relatively homogenous disease that has readily distinguishable hereditary and sporadic forms. The two-hit hypothesis developed by Knudson provided the first model for understanding cancer predisposition (see Chapter 3). In comparison with retinoblastomas, colorectal tumors are plentiful and can be found in individuals at different ages. Because tumor samples can be obtained during routine colonoscopy, all stages of growth have been subject to detailed analysis. The multistage model of tumorigenesis that emerged from studies of colorectal cancers provides a paradigm for understanding how cancer clones evolve and expand.

F. Bunz, *Principles of Cancer Genetics*.

© Springer 2008

Most cancers are not understood at the same level of detail as retinoblastoma and colorectal cancer. The reasons that many cancers remain incompletely characterized include clinical heterogeneity, access by geneticists to insufficient numbers of clinical samples representing different stages of disease, and a lack of a clearly diagnosable hereditary form of the disease that allows the mapping of a predominant gatekeeper gene and pathway. Despite these obstacles to progress, cancer genes have been found in all of the most common types of cancer. The wider application of large-scale sequencing approaches to cancer genomes promises to reveal many more cancer genes in the near future.

Cancer Incidence and Prevalence

According to statistics compiled by the National Cancer Institute, it is projected that over 1.4 million people will be diagnosed with cancer in the USA, and over 550,000 people will die of cancer, yearly. Among these dire figures, a moderate but

Table 6.1 The most commonly diagnosed cancers in the US population in 2007, by incidence rate. Data are from the Surveillance, Epidemiology, and End Results Program of the National Cancer Institute. In the case of cancers that occur in only one sex (e.g. prostate cancer in men) only the sex-specific population is considered. Excluded are nonmelanoma skin cancers, which are very common, but rarely lethal

Cancer site	Incidence rate (per 100,000)	Age at diagnosis	Mortality rate (per 100,000)	Lifetime risk[1] (%)	Trend[2]
Prostate	168	68	27.9	17	↔
Breast	127.8	61	25.5	12	▼
Lung	64.5	70	54.7	7.0	▼
Colon and rectum	51.6	71	19.4	5.4	▼
Uterine endometrium	23.2	63	4.1	2.5	▼
Lymphoma	22.0	64	8.1	2.4	↔
Bladder	21.1	73	4.3	2.4	▼
Melanoma of the skin	18.5	59	2.6	1.7	▲
Ovary	13.5	63	8.9	1.4	▼
Kidney	12.8	71	4.2	1.3	▲
Leukemia	12.3	67	7.5	1.3	▼
Pancreas	11.4	72	10.6	1.3	↔
Oral cavity and pharynx	10.5	62	2.7	1.0	▼
Uterine cervix	8.7	48	2.6	0.7	▼
Thyroid	8.5	47	0.5	0.7	▲
Stomach	8.1	74	4.2	0.9	▼
Brain and nervous system	6.4	55	4.4	0.6	▼
Liver	6.2	65	4.9	0.7	▲

[1] Based on rates from 2002 to 2004, the percentage of people born today who will be diagnosed with cancer during their lifetime.

[2] Statistically significant change in annual incidence, during the most recent period for which cumulative data are available.

significant downward trend in the rates of cancer incidence and mortality over the last 10 years has provided some encouragement.

An *incidence rate* of a specific cancer is defined as the number of newly diagnosed cases that will occur during a given year. This number may include multiple primary cancers that occur in a single patient, but does not typically include recurrences after treatment. The mortality rate is simply the rate of deaths that are directly attributed to a given cancer. Cancer incidence and mortality are most often expressed as rates per 100,000 people at risk. The most lethal forms of cancer are those in which the mortality rate approaches the incidence rate (see Table 6.1). Another term frequently used to describe the impact of cancer upon a population is the *prevalence*. Cancer prevalence is defined as the population that, on a given date, has previously had a diagnosis of a cancer. The prevalence therefore includes both new and preexisting cases of a cancer, and is a function of both past incidence and survival.

Many cancers show strikingly different rates of incidence and mortality in different ethnic populations. In some cases, genetic factors may affect cancer predisposition within some groups. Economic and social status can combine to increase or decrease the lifetime risk and confound genetic analysis. Access to healthcare and screening services and exposure to carcinogens related to diet and lifestyle are but a few of the many non-genetic factors that are known to strongly influence both incidence and mortality.

Lung Cancer

Lung cancer is the leading cause of cancer death in the USA. More than 90% of lung cancers develop as a direct result of exposure to tobacco smoke. Approximately 10% of smokers eventually develop lung cancer.

Tumors in the lung arise from epithelial cells that line the alveoli, bronchioles and bronchi (see Fig. 6.1). There are four histologic types of lung cancer that fall into two broad treatment groups. Squamous cell, adeno- and large-cell carcinomas are collectively referred to as the non-small-cell lung carcinomas (NSCLC), which together compose 75% of lung tumors. Because this group of cancers tends to metastasize at a later point in the disease, early detection and surgical resection result in many cures. The remaining one quarter of lung cancers are small cell lung carcinomas (SCLC), the most aggressively metastatic tumors and therefore the most difficult to effectively treat. Cigarette smoking has been conclusively shown to be causally related to both groups of lung cancers.

Unlike many of the most common forms of cancer, lung cancer does not occur in a classical familial form. Therefore, there is no obvious gatekeeper gene that is known to strongly affect predisposition. Nonetheless, there is ample evidence that genetic factors do influence the incidence of lung cancer in at-risk smokers. For example, patients diagnosed with retinoblastoma have been found to have an

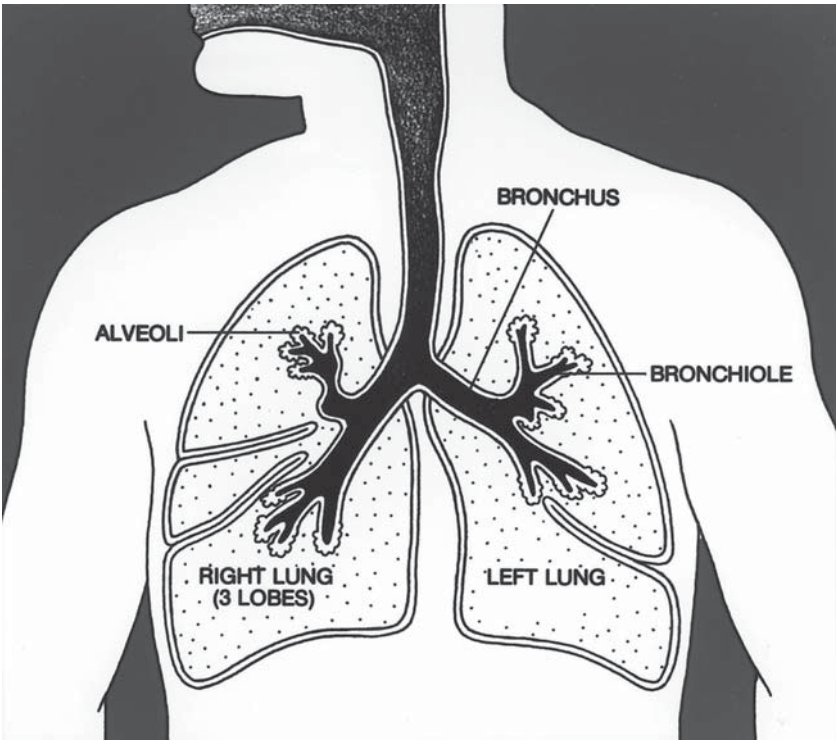

Fig. 6.1 Anatomy of the lungs. Large airways become smaller as they progressively branch, ultimately terminating in the alveoli. (Courtesy of the National Cancer Institute.)

increased incidence of lung cancer later in life compared with the normal population. Thus, while the penetrance of *RB* mutations with respect to the development of retinoblastoma is nearly 100%, the penetrance of *RB* mutations with respect to lung cancer is much lower but still significant.

Several cancer genes have been found at high frequency in sporadic lung cancers. The gene known to be mutated most frequently in lung cancers is *P53. P53* mutations are found in about 50% of NSCLC and in over 90% of SCLC. As described in Chapter 1, many smoking-associated mutations are G→T transversions that occur in known hotspots of the *P53* open reading frame. These characteristic mutations can be directly attributed to bulky adducts caused by exposure to BPDE, a carcinogen in cigarette smoke.

As with the germline mutations in *RB* that play a definite role in lung cancer predisposition, somatic *RB* mutations are found in a significant proportion of sporadic lung cancers. *RB* is inactivated in 30–40% of NSCLC and in nearly all SCLC tumors. Among NSCLC, *RB* mutations are associated with more advanced tumors, implying that *RB* loss occurs during later stages of tumorigenesis. Deletions

affecting the gene that encodes the cell cycle regulator p16 are also frequently found in NSCLC. Because SCLC frequently inactivate *RB*, there would be little selective pressure to also inactivate the p16 gene, *CDKN2A*, which functions in the same pathway (see Chapter 5). Indeed, *CDKN2A* mutations in SCLC have not been reported. Similarly, a gene that encodes cyclin D, *CCND1*, is amplified in a proportion of NSCLC, but not in SCLC.

The receptor tyrosine kinase pathway is frequently altered in the predominant subtype of NSCLC, the lung adenocarcinoma. Collectively, mutations in *EGFR, ERBB2, PIK3CA, BRAF* and *K-RAS* that constitutively activate EGFR signaling have been observed in 50% of lung adenocarcinomas. A specific germline mutation in EGFR, T790M, has been associated with an inherited susceptibility to lung adenocarcinomas.

Mutations in *K-RAS*, as well as less frequent amplifications of *C-MYC*, have been shown to have prognostic significance. Amplification of *PIK3CA*, and the attendant dysregulation of the PI3K pathway, is a common occurrence in NSCLC of the squamous cell subtype.

Prostate Cancer

Prostate cancer is the most commonly diagnosed cancer in men. Over 35% of all cancers affecting men are prostate cancers. The high incidence of this disease has contributed to a prevalence that is over 2 million in the USA. To a greater extent than most cancers, prostate cancer is strongly associated with aging and thus extremely rare in men below the age of 40.

The prostate is a walnut-sized gland located near the base of the urinary bladder (see Fig. 6.2). The majority of the lesions that develop into prostate malignancies arise in the periphery of the gland, while approximately 20% of premalignant lesions arise occur in the region that surrounds the urethra, known as the *transition zone*. The transition zone frequently undergoes hypertrophy, causing a common condition known as benign prostatic hyperplasia (BPH). BPH is not believed to be a premalignant condition.

The initiation of tumors in the prostate gland is a very frequent event. Nearly one third of all men over the age of 45 have histologically identifiable prostate cancer precursor lesions known as *prostatic intraepithelial neoplasia* (PIN). Many PIN lesions are multifocal (see Fig. 6.3), suggesting that they represent multiple lesions that arose independently. PIN lesions are thought to be the precursors of prostate cancer, but the majority of PIN lesions do not progress to clinically detectable prostate tumors.

The evolution of prostate tumors clearly has a highly variable course. While the rate of PIN development is similar throughout the world, the rates of prostate cancers in different populations are highly dependent on race and geographic location. This variability is probably due to a combination of genetic and environmental factors, most of which remain unidentified.

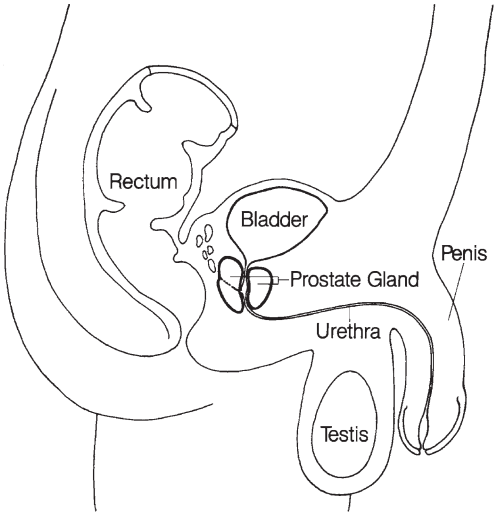

Fig. 6.2 Location of the prostate gland. Positioned near the base of the urinary bladder, the prostate encircles the urethra. Tumors or benign hypertrophy can cause urinary obstruction. (Courtesy of the National Cancer Institute.)

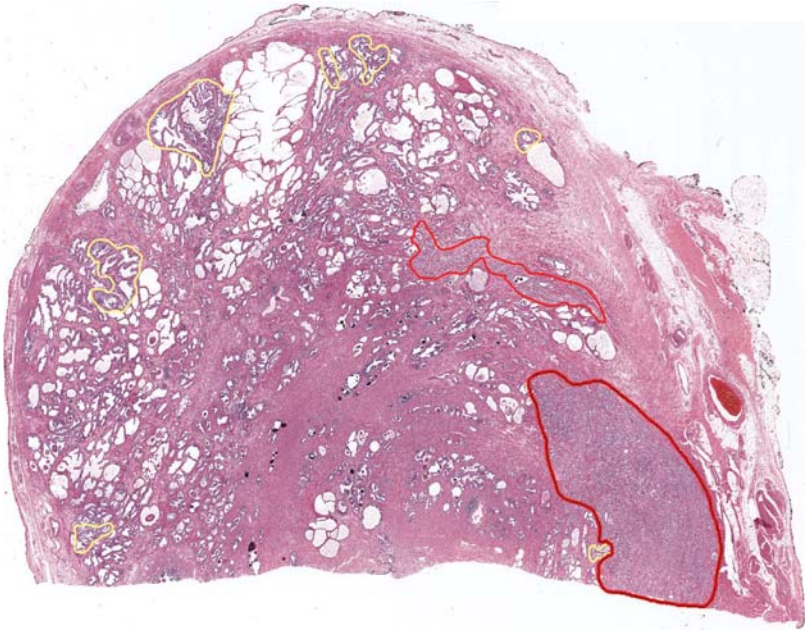

Fig. 6.3 Multifocal neoplasia in the prostate. Prostate intraepithelial neoplasia (PIN; circled in yellow) and prostate cancer (circled in red) can be seen in a single section of prostate tissue viewed under low magnification. (Courtesy of Angelo De Marzo, M.D., Ph.D., Johns Hopkins University.)

Also variable is the genetic etiology of prostate cancer, which varies significantly from case to case. DNA sequence analysis has revealed tumor-associated mutations in a number of known cancer genes, including *P53, PTEN, RB* and *CDKN2A. P53, PTEN* and *CDKN2A* mutations are more commonly found in metastatic tumors, suggesting that loss of these tumor suppressors occurs during the later stages of cancer progression. Unlike the case of other common cancers, no single gene has been found to be mutated in the majority of prostate cancers.

Cases of prostate cancer have been found to cluster in high-risk families. Studies of familial aggregation suggest that 5–10% of prostate cancers are attributable to the inheritance of autosomal dominant cancer genes. Men that have a family history of the disease reportedly have a relative risk of 3.3 for prostate cancer, as do men that carry germline mutations in *BRCA1*, which are more commonly associated with breast cancer in women. While such evidence suggests that prostate cancer has a significant hereditary component, alleles that strongly predispose carriers to prostate cancer have not been identified.

Analysis of sporadic prostate cancers has revealed several recurrent chromosomal alterations. The most frequent of these are gains and losses of sequences on two regions on the short arm of chromosome 8. Deletions affecting 8p21 are found in two thirds of all PIN lesions, while 8p22 is deleted in most adenocarcinomas. Amplification events affecting the *C-MYC* locus and many other genes on the long arm of chromosome 8 frequently cause gains of sequence. Similarly, frequent alterations have been found to affect defined regions of chromosomes 5, 6, 7, 10, 13, 16, 17, and 18. Presumably, tumor suppressor genes (in the lost regions) and oncogenes (in the gained regions) provide a selective advantage for these chromosomal changes, but these putative cancer genes remain largely unidentified.

Breast Cancer

Breast cancer is the most commonly diagnosed cancer – and the leading cause of cancer mortality – in women. Over 2.4 million women in the USA have a history of breast cancer, of whom approximately 30% will ultimately die of the disease. After increasing over a period of several decades, the rates of breast cancer incidence and mortality have moderately declined since 2001.

The majority of breast cancers arise from the epithelia that line the milk-producing lobules and ducts of the mammary gland (see Fig. 6.4). All women have a similar number of these cells, regardless of overall breast size. For this reason, breast size is not a significant risk factor for breast cancer. Approximately 80% of breast cancers are ductal in origin, between 5–10% are infiltrating lobular carcinomas and the remainder arise from diverse cell types.

Like several common types of cancer, breast cancer begins with defined precursor lesions. Small, noninvasive lesions, typified by *ductal carcinomas in situ* (DCIS), are believed to progress to first invasive and then metastatic lesions. In contrast, small lesions found in the lobular epithelia, known as *lobular carcinomas in situ*

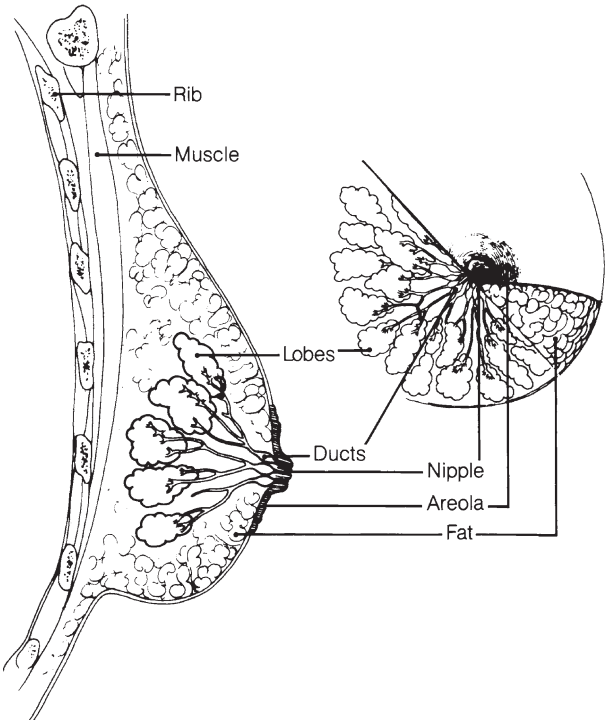

Fig. 6.4 Anatomy of the female breast. Milk produced in the lobules is transported to the nipple by the breast ducts. The bulk of breast mass is composed of stromal and fatty tissue. (Courtesy of the National Cancer Institute.)

(LCIS), are not believed to be precursor lesions that progress, although their appearance is associated with subsequent disease.

There are many risk factors for the development of breast cancer, the single most significant of which is a positive family history. About 5% of breast cancers are thought to be hereditary in nature, arising as a consequence of germline alleles that confer cancer predisposition. Two high-penetrance tumor suppressor genes have been described, *BRCA1* and *BRCA2* (see Chapter 3). Together, these two genes account for a significant proportion of familial breast cancers. Mutations in *BRCA1* have been identified in 15–20% of women with a family history of breast cancers. The median age of breast cancer onset in *BRCA1* mutation carriers is 42 years, which is 20 years younger than what is observed in sporadic cases.

BRCA1 and BRCA2 proteins function in pathways involved in DNA repair and cell-cycle regulation (see Chapter 5). Other DNA repair proteins that functionally interact with BRCA1 and BRCA2, such as Chk2 and the FANC gene-encoded proteins, are candidates for low-penetrance breast cancer genes. Breast cancer is also part of the clinical spectrum of Li Fraumeni syndrome, Cowden disease and ataxia telangiectasia, demonstrating that germline mutations in *P53, PTEN* and *ATM*, respectively, confer a significant risk (see Chapter 4).

Several genes that are widely found to be mutated in other common cancers are also mutated in sporadic breast cancers. Nearly 30% of breast cancers harbor *P53* mutations. More that 20% of breast cancers harbor somatically acquired mutations in *PIK3CA* and approximately 9% exhibit *PIK3CA* amplification. Other genes that are frequently amplified in breast cancers include *CMYC* and *ERBB2* (also known as *HER2/neu*).

Analysis of global gene expression in breast cancers has revealed numerous genes that are frequently overexpressed. *C-SRC, CCND1* (a gene that encodes cyclin D), and *BCL2* are examples of genes that have been reported to be expressed at high levels in many breast cancers. It is important to note that in most cases, the genetic alteration underlying the expression defect has not been determined. It therefore remains a possibility that overexpression of some genes in breast cancers is an indirect effect caused by dysregulated upstream pathways, rather than a mutation at the locus in question.

As is the case in prostate cancer, many recurrent chromosome abnormalities have been described in breast cancers. Common regions of LOH have been described on chromosomes 1, 3, 6, 7, 11, 13, 16, 17, and 18. The majority of the genes that underlie the apparent selection of these alterations have not been identified.

Endometrial Cancer

Endometrial cancer is the most common malignancy of the female genital tract and the fourth leading cause of cancer death in women in the USA. Though the majority of cases appear to be sporadic, a significant number have a known, heritable etiology. Endometrial cancer is the most common extracolonic cancer in the most prevalent cancer predisposition syndrome, hereditary nonpolyposis colorectal cancer (HNPCC; see Chapter 4). Women who carry HNPCC alleles have a relative risk of endometrial cancer that is tenfold higher than the general population.

The endometrium that lines the interior of the uterus (see Fig. 6.5) is composed of both epithelial and stromal cells. While cancers can arise from both of these cell types, more than 95% of endometrial cancers are carcinomas arising from the epithelia. Endometrial carcinomas, can be further categorized by histological criteria into two subtypes: endometroid carcinoma and uterine serous carcinoma. Endometrioid carcinoma arises in a stepwise manner from a noninvasive precursor lesion called *complex atypical hyperplasia* (CAH). The less-common uterine serous carcinoma develops in the setting of epithelial atrophy, from a precursor lesion called *endometrial intraepithelial carcinoma* (EIC). These lesions are distinguishable both clinically and at the molecular level.

Endometrial cancer is a well-defined component of HNPCC. Of sporadic cancers, approximately 25% exhibit microsatellite instability that is indicative of defective DNA mismatch repair (see Chapter 4). This defect is found in both endometrioid carcinomas as well as in its precursor lesion, the CAH.

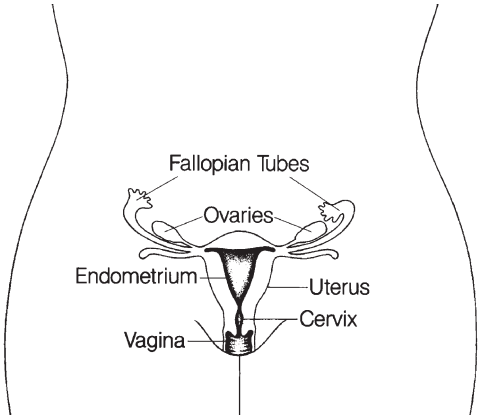

Fig. 6.5 Anatomy of the female pelvis. Common cancers arise from several organs in the female reproductive tract, including the ovaries, the endometrium that lines the uterus and the uterine cervix. Rare cancers affect the vagina and fallopian tubes. (Courtesy of the National Cancer Institute.)

The most common mutations in endometrial cancers affect the PI3K pathway. Approximately 50% of endometrioid carcinomas have mutations in *PTEN. PTEN* mutations appear to occur early in tumorigenesis, as they are also found in a significant proportion of CAH. Many *PTEN* mutations occur in a region that encodes the phosphatase domain, and result in loss of protein expression. *PIK3CA* mutations are found in approximately 40% of endometrioid carcinomas, often coexisting with a *PTEN* mutation. As described in Chapter 5, mutations in one component of a pathway often preclude the selection of other genes that populate the same pathway. Mutations of *PTEN* and *PIK3CA* appear to violate this general principle. Experimental evidence has demonstrated that reduction of PTEN activity in the context of a *PIK3CA* mutation causes enhanced phosphoryation of AKT. Therefore, it appears that coincident mutation of *PTEN* and *PIK3CA* leads to hyperactivation of PI3K signaling (see Fig. 6.6). Unlike *PTEN* mutations, mutations in *PIK3CA* are not commonly found in precursor lesions, suggesting that activation of *PIK3CA* occurs upon tumor invasion.

Mutations in *K-RAS, P53*, and *CTNNB1* are also fairly common in endometrioid carcinomas. These genes demonstrate an interesting overlap between mutations found in colorectal cancers and in the endometrioid form of endometrial cancer.

The two histologically distinguishable forms of endometrial carcinoma show a molecular etiology that is also distinct. *PTEN* mutations are rare in the less-common serous carcinoma, while *P53* mutations occur at a frequency of greater than 90%. Microsatellite instability is very uncommon in serous carcinomas.

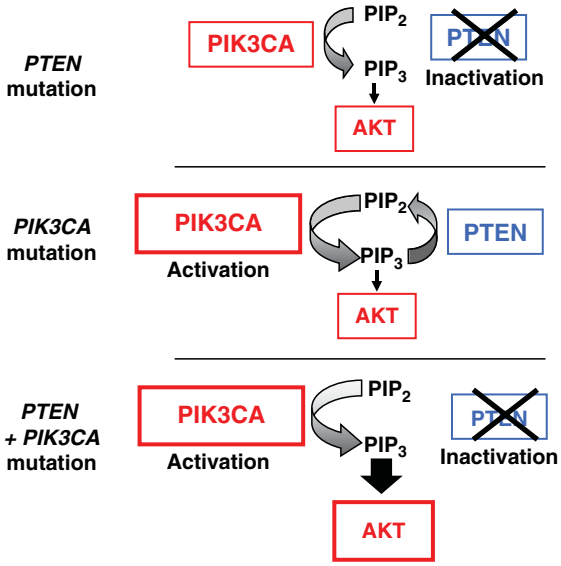

Fig. 6.6 Coincident mutations in *PTEN* and *PIK3CA* in endometrial cancers. Mutations that inactivate *PTEN* (top panel) or activate *PIK3CA* (middle panel) cause ligand-independent activation of AKT. Mutation of both *PTEN* and *PIK3CA* leads to hyperphosphorylation of AKT and increased PI3K pathway activation. Mutational data suggest that normal regulation of PI3K signaling is a critical element of homeostasis in the endometrial epithelia.

Lymphoma

A large and diverse group of malignancies are derived from lymphocytes and their precursors. The lymphoid malignancies that grow as solid tumors are known as lymphomas. This is a bimodal incidence of lymphoma with respect to age at diagnosis. Hodgkin lymphoma is a disease of young adulthood with a median age at diagnosis of 38. Approximately 12% of cases occur in individuals less than 20 years of age. A second peak of Hodgkin lymphoma occurs later in life. The most prevalent lymphomas are non-Hodgkin lymphomas, which are diagnosed at an average age of 67.

The non-Hodgkin lymphomas can be subdivided by histological and anatomical criteria into numerous subtypes that include (in order of incidence) diffuse large B-cell lymphoma, follicular lymphomas, muscosa-associated lymphoid tissue (MALT) lymphoma, mantle cell lymphoma and Burkitt lymphoma. Depending on the type, tumors can occur in lymph nodes and at a variety of extranodal sites.

The most obvious genetic defects in lymphoma cells are chromosomal translocations. These genetic alterations arise as a consequence of a high rate of gene recombination that is intrinsically related to the ontogeny of cells in the lymphoid lineage. Presumably, genes near the common breakpoints are dysregulated as a

result of their translocation. Many of the genes that drive selection of these recurrent translocations remain unknown. In several cases, the genes that have been identified are unique to lymphoid cancers and are not altered in other tumor types.

Among the genes identified at common breakpoints in follicular lymphomas is *BCL2*, an antiapoptotic gene on chromosome 18 that is fused to the immunoglobulin heavy chain gene promoter. Approximately 40% of diffuse large-cell lymphomas and 5–10% of follicular lymphomas harbor translocations that dysregulate expression of BCL6, a DNA binding protein that represses transcription of specific target genes. In MALT lymphomas, a gene designated *MALT1*, located 5 Mb centromeric to *BCL2*, is frequently fused to a highly expressed inhibitor of apoptosis, *API2*, on chromosome 11. *MALT1* is required for a variety of inflammatory processes that involve the activation of the pro-proliferative transcription factor NF-κB. MALT lymphomas that occur in the stomach are associated with infection with *Helicobacter pylori*, indicating the role of chronic inflammation in this disease subtype (see Chapter 1). Mantle cell lymphomas have a recurrent translocation between chromosomes 11 and 14 that results in overexpression of *CCND1* (initially designated BCL1), the gene that encodes Cyclin D1.

In addition to genes located at translocation breakpoints, a number of genes have been found to be somatically mutated by other mechanisms. *P53* is inactivated by point mutations in approximately 20% of non-Hodgkin lymphomas. In a small number of cases, constitutive activation of the RAS signaling pathway results from mutations in *BRAF* or *K-RAS*.

Burkitt lymphomas are relatively rare outside of Africa, where they are highly endemic. This aggressive form of non-Hodgkin lymphoma is always associated with chromosomal translocations that result in the overexpression of *C-MYC*. In addition, approximately 30% of Burkitt lymphomas harbor *P53* mutations. Infection with Epstein Barr virus (EBV) is strongly associated with Burkitt lymphoma, but the role of viral genes in tumorigenesis remains unclear.

The second major category of lymphomas, Hodgkin lymphomas, feature recurrent gains of sequence on chromosomes 2, 9 and 12, and respective amplification of the oncogenes *REL* (the cellular homolog of the *V-REL* viral oncogene), *JAK2* (which encodes a tyrosine kinase), and *MDM2*. Infectious mononucleosis caused by EBV is a significant risk factor for Hodgkin lymphoma, and EBV DNA sequences have been found in a significant proportion of Hodgkin lymphoma biopsies. As is the case with Burkitt lymphoma, the role of EBV in Hodgkin lymphoma progression is currently unknown.

A small proportion of both non-Hodgkin lymphomas and Hodgkin lymphomas appear to represent familial forms of the disease. The underlying basis for lymphoma predisposition remains unknown.

Bladder Cancer

Cancer of the urinary bladder has a fourfold higher incidence in males, in whom it is the fourth most common malignancy. Bladder cancer is strongly associated with local irritation and inflammation caused by environmental toxins in the

urine. In the USA, cigarette smoking is the most significant risk factor, contributing to 45% of cases.

Almost all bladder cancers arise from the urothelium, a specialized type of epithelium (also known as transitional epithelium) that lines the urinary tract (see Fig. 6.7). There are two distinct early forms of transitional carcinomas: carcinoma in situ and papillary tumors. Carcinoma in situ is a flat lesion with a high propensity for progression. It has been demonstrated that 40% of patients with carcinoma in situ progress to invasive disease within 5 years. Papillary lesions tend to recur after treatment, but have a less than 20% risk of progression.

Many bladder cancers appear multifocal at the time of presentation. Alone, this finding would suggest the simultaneous occurrence of multiple independent lesions. Contrary to this interpretation, molecular analyses have revealed that multiple foci are in fact all clonally derived from a single progenitor cell. It would therefore appear that early neoplastic cells within the urothelium are highly mobile.

The most frequently observed genetic alterations in bladder cancers are chromosomal deletions. Complete loss of chromosome 9 and resultant monosomy is the most common cytogenetic abnormality. Deletions of sequences on chromosome 9p that inactivate *CDKN2A* occur at early stage of both flat carcinoma in situ and in papillary lesions. At least 50% of all bladder cancers contain a deletion that results in loss of p16 expression. Deletions on 17p and 13q that inactivate *P53* and *RB*, respectively, have been found to occur frequently during later, invasive stages of the disease.

Several proto-oncogenes have been found to be activated in bladder cancers. The RAS gene family was first identified in a bladder cancer cell line (see Chapter 2).

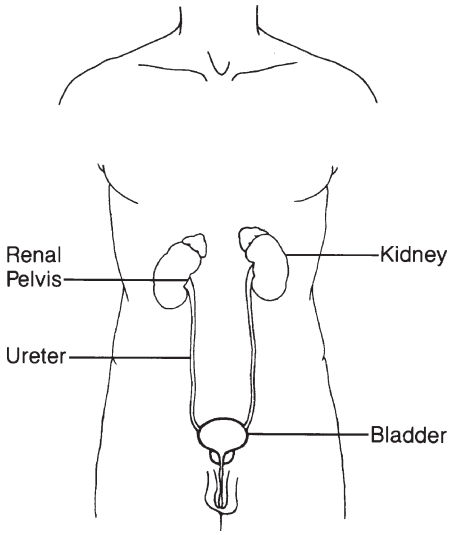

Fig. 6.7 Anatomy of the male urinary tract. The male urinary tract is composed of the kidneys, ureters, the urinary bladder and the prostate. (Courtesy of the National Cancer Institute.)

Mutation in the RAS family member *H-RAS* have been variously reported at frequencies ranging between 6% and 44%. Amplification of a gene designated *CDC91L1* has been found in greater than 30% of tumors while *ERBB2* and *CCND1* amplifications are found in fewer than 20% of tumors.

While the majority of bladder cancers are sporadic, a small number of inherited cases occur in the context of HNPCC. Bladder cancer is the fourth most common malignancy in HNPCC patients. Defects in mismatch repair also contribute to sporadic cases, of which approximately 2% exhibit evidence of microsatellite instability.

Melanoma of the Skin

Melanoma is a common, deadly form of skin cancer that arises from melanocytes, the pigment producing cells in the skin. The major risk factor for melanoma is exposure to the ultraviolet (UV) component of sunlight. The incidence of melanoma in the USA has markedly increased over the past several decades, particularly in the southern latitudes, mirroring an increase in the popularity of tanning and outdoor activities.

Melanocytes arise from embryonic precursors that migrate from the central nervous system into the skin during development. The migratory nature of the melanocyte lineage may in part explain the extreme extent to which melanomas tend to spread and metastasize.

While melanoma develops in individuals from all ethnic groups, the incidence of melanoma is significantly higher among light-skinned individuals. This increased risk is not the result of a larger number of target cells to be mutated, as light-skinned and dark-skinned individuals have a similar number of melanocytes. Rather, each melanocyte in a dark-skinned individual produces more pigment that confers protection against ultraviolet radiation.

In addition to environmental factors, genetic factors also play a large part in overall risk. While the majority of melanomas are sporadic, approximately 10% of cases occur in high-risk families, including those affected by the Familial Atypical Multiple Mole Melanoma (FAMMM) syndrome. These families feature a high incidence of pigmented lesions known as atypical nevi (see Fig. 6.8).

The most common cytogenetic aberration in melanoma cells is deletion within the short arm of chromosome 9, in the region of *CDKN2A*. Inactivating germline mutations in *CDKN2A* are a cause of FAMMM syndrome. The penetrance of inherited mutant alleles is nearly 70%. Mutations in *CDKN2A* are also found in sporadic melanoma cases at a frequency of approximately 20%. Mutations that are somatically acquired often exhibit the UV signature (see Chapter 1), including C→T or CC→TT transitions.

Genetic evidence suggests an important role for the regulation of the $G_1 \rightarrow S$ cell cycle transition in the suppression of melanoma (see Fig. 6.9). The *RB* gene product functions in a common pathway with p16, the product of the *CDKN2A* gene (see Chapter 5). Carriers of germline *RB* mutations who are successfully treated for retinoblastoma in childhood are at an 80-fold risk of developing melanoma later in

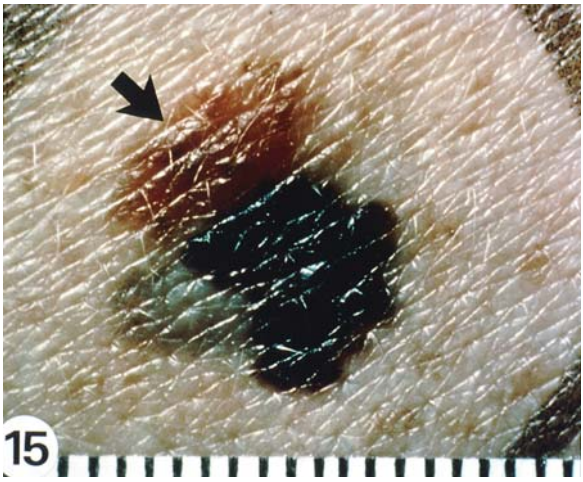

Fig. 6.8 A melanoma arising from a dysplastic nevus. The 4-by-8-mm, pink-tan lesion with irregular borders at the upper left (arrow) is a dysplastic nevus. Arising from this lesion is an invasive malignant melanoma, with its characteristic blue–black color, notched border, and distorted surface. (Courtesy of the National Cancer Institute.)

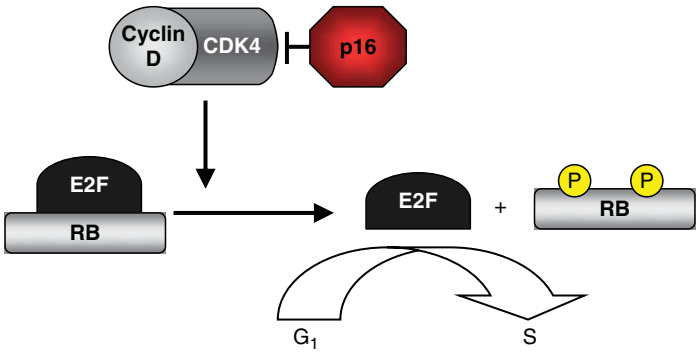

Fig. 6.9 Interactions between RB, CDK4 and p16 in melanoma tumor suppression. The phosphorylation of RB by activated CDK4 controls the transition from $G_1 \rightarrow S$ phase of the cell cycle. This reaction is blocked by the cyclin dependent kinase inhibitor p16. Melanomas can arise as a result of loss-of-function mutations in RB or *CDKN2A* (which encodes p16), or gain-of-function mutations in the proto-oncogene *CDK4*.

life. *RB* mutations have been found in sporadic lesions as well. Another gene that affects cell cycle regulation in concert with *CDKN2A* and *RB* is *CDK4*, which encodes a cyclin dependent kinase. CDK4 activity, which directly controls the $G_1 \rightarrow$ S cell cycle transition, is directly inhibited by p16. Analysis of melanoma-prone families without *CDKN2A* mutations has revealed mutations that disrupt the p16 binding site of the *CDK4* encoded protein. As in the case of *CDKN2A, CDK4*

mutant alleles are highly penetrant. Thus, *CDK4* is another rare example of an oncogene that confers predisposition to cancer when inherited in the germline.

The most frequent genetic alterations in sporadic melanomas are mutations in *BRAF*, which functions in the RAS pathway. Activating mutations in *BRAF* have been found in pigmented nevi, as well as in more than 50% of localized and metastatic melanomas. As in other cancers with *BRAF* mutations, the most common mutated allele is V600E (see Chapter 5). While the genetic alteration that underlies the V600E codon change is not a typical UV signature mutation, the *BRAF* V600E allele is found most commonly in melanomas that occur in sun-exposed areas. In a significant number of melanomas, the RAS pathway is constitutively activated by *N-RAS* and, less frequently, *K-RAS* mutations.

There are several other cancer gene pathways that appear to be disrupted during the development of melanomas. Losses of sequence on chromosome 10q occurs early and frequently during melanoma tumorigenesis. *PTEN*, which maps to this region has been reported to be inactivated by both deletions and point mutations in a significant fraction of tumors. Activation of receptor tyrosine kinases, including EGFR, C-MET and C-KIT, has been reported in melanomas; the genetic alterations that might underlie these activities have generally not been determined. Interestingly, *P53* mutations are rarely found in melanomas, unlike the other common cancers.

Ovarian Cancer

Ovarian cancer is the fifth leading cause of cancer deaths among women in the USA. Early tumors are rarely detected; most ovarian cancers are spread throughout the pelvis at the time of diagnosis. Most ovarian cancers are sporadic, but a significant number are associated with known familial cancer syndromes. A major risk factor is cumulative ovulatory activity, which is often increased by early onset of menarche and nulliparity.

Ovarian cancers are a heterogeneous group of histologically distinct tumors. The most common ovarian cancers are ovarian epithelial carcinomas that begin in the cells on the surface of the ovary. Histological subtypes include serous, clear cell, endometrioid and mucinous forms. While many of these subtypes bear marked similarity to other tumors of the female genital tract, the precise cellular origin of many ovarian cancers remains undetermined.

The subtypes of ovarian tumors arise from distinct precursor lesions and generally exhibit distinct genetic alterations. Molecular analysis has shown that low- and high-grade serous carcinomas probably arise via alterations in different pathways, the former involving mutations in *K-RAS* and *BRAF* and the latter involving mutations of *P53*. *P53* is mutated in approximately 60% of serous and endometrioid carcinomas, but less often in mucinous and clear cell carcinomas. Mucinous ovarian tumors, which probably arise from precursor adenomas, exhibit a high frequency of *K-RAS* mutations. Endometrioid carcinomas arise from endometriosis, an inflammatory lesion that resembles the lining of the uterine endometrium. These

cancers harbor frequent mutations in *CTNNB1* that constitutively activate the WNT/APC pathway. Endometrioid tumors with dysregulated WNT/APC signaling frequently have concurrent mutations that dysregulate PI3K signaling, most often in *PTEN* but also in *PIK3CA*.

Many ovarian cancers exhibit evidence of gene amplification. Several genes that appear to drive the selection of frequently amplified regions have been identified, including *CCNE1* (which encodes Cyclin E1), *ERBB2, AKT2, PIK3CA* and the MYC family member *L-MYC*.

Up to 10% of all ovarian cancers are estimated to occur in individuals with an inherited predisposition. Familial ovarian cancer is a significant component of two major syndromes: (1) familial breast cancer (associated with germline mutations in *BRCA1* and *BRCA2*; see Chapter 3) and (2) HNPCC (associated with alterations in mismatch repair genes; see Chapter 4).

Cancer of the Kidney

The predominant form cancer that occurs in the kidney, renal carcinoma, accounts for approximately 3% of all adult cancers in the USA. Males are affected twice as frequently as females, and individuals in end-stage renal failure have a risk of cancer that is up to 30-fold that of the general population. Environmental risk factors include cigarette smoking and exposure to asbestos. An estimated 4% of all renal cell carcinomas are hereditary in origin.

Renal carcinoma occurs in several histological types, the most common of which (85%) features a type of cell known as the clear cell. The remaining 5–10% of tumors exhibit a papillary morphology. Both of these tumor types occur in sporadic and hereditary cases.

There are three types of hereditary renal carcinoma: (1) von Hippel-Lindau (VHL), (2) hereditary clear cell renal carcinoma, and (3) hereditary renal papillary cell carcinoma. Similar to the pattern of retinoblastoma, sporadic kidney cancers are typically solitary lesions, while hereditary disease is often multifocal and bilateral.

The best understood hereditary kidney cancer is VHL. VHL is caused by germline mutations in *VHL*, a 3-exon gene located at 3p26-p25 and cloned in 1993. A variety of inactivating mutations have been found throughout the *VHL* open reading frame, including small insertions and deletions and single base substitutions. *VHL* is a tumor suppressor gene that has all of the characteristics of a gatekeeper (see Chapter 4). Germline mutations in *VHL* are highly penetrant. Loss of heterozygosity (LOH) during tumorigenesis completes the inactivation of both alleles. VHL is also inactivated in the majority of sporadic clear cell renal carcinomas, at very early stages of the disease.

Clinically, VHL is a heterogeneous disorder. Affected individuals are at elevated risk for development of uncommon tumors that affect the kidney, cerebellum, spine, eye, inner ear, adrenal gland and the pancreas. While *VHL* mutations have been found in sporadic forms of these types of tumors, *VHL* mutations are not found in the more common cancers.

Hereditary papillary cell carcinoma is caused by germline mutations that activate the *MET* proto-oncogene (see Chapter 2). *MET* encodes a receptor tyrosine kinase. Hotspots for mutations in *MET* are common to those in related receptor tyrosine kinases such as *RET*, which is mutated in the germline of individuals affected by Multiple Endocrine Neoplasia type 2.

Leukemia

The leukemias are a diverse group of cancers that affect the cells of the blood and blood-forming tissues. Over 200,000 people in the USA have a current or past diagnosis of leukemia. While leukemia is the leading form of pediatric cancer, leukemia affects many more adults than children.

Leukemias arise in blood-forming tissues such as the bone marrow, from cells of the lymphoid or the myeloid lineage. Both lymphocytic and myeloid leukemias can present in acute and chronic forms. Acute myeloid leukemia (AML; see Fig. 6.10) and chronic lymphocytic leukemia (CLL) are the most common leukemias in adults; acute lymphocytic leukemia (ALL) is the most common pediatric cancer. Additional subcategories of leukemias arise from cells at various stages of hematological development.

Chromosomal translocations are found in over 50% of leukemias in both in children and adults. As described in Chapter 2, translocations can activate proto-oncogenes near breakpoints in two ways: (1) by fusing together the coding sequences of two genes that are normally unrelated, or (2) by placing a gene under the transcriptional control of an unrelated gene that is expressed at high levels.

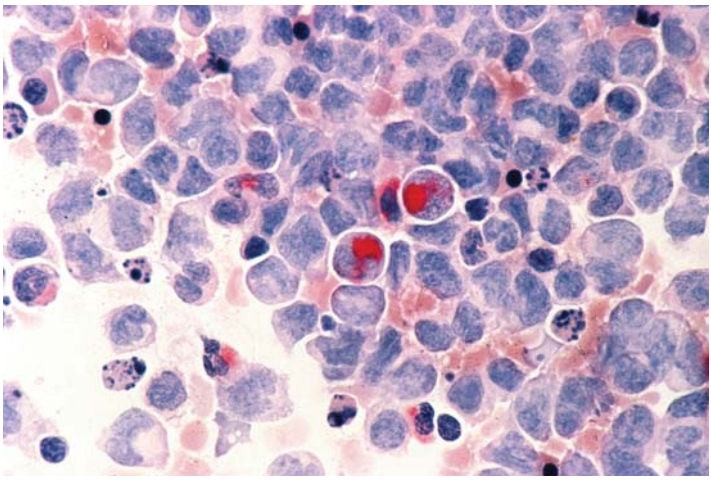

Fig. 6.10 Acute myelocytic leukemia cells. AML cells obtained from cardiac fluid, stained with esterase, at 400X magnification. (Courtesy of the National Cancer Institute.)

The classic example of oncogene activation by translocation is the translocation between chromosomes 9 and 22 that creates the Philadelphia chromosome, originally observed in chronic myelocytic leukemia (CML), and subsequently detected in ALL (see Chapter 2). This translocation event creates a hybrid gene that contains downstream elements of *ABL* (which encodes a tyrosine kinase) fused with upstream elements of *BCR* (a gene that is highly expressed).

Recurrent translocations involving chromosome 11 have been observed in aggressive forms of both AML and ALL. The gene at the common breakpoint of these translocations is designated *MLL*, for mixed lineage leukemia. Different translocation events activate *MLL* activity by creating in-frame fusion proteins. *MLL* encodes a protein with intrinsic methyltransferase activity that, by catalyzing the methylation of cytosine bases (see Chapter 1), regulates the expression of known mediators of development.

Another common breakpoint gene is *TCL1*, for T-cell leukemia. Translocations involving chromosome 14q frequently place the *TCL1* locus in the vicinity of a highly expressed T-cell receptor gene. The protein encoded by *TCL1* has been shown to directly bind AKT within its pleckstrin homology domain, and thus activate the PI3K/AKT pathway (see Chapter 5).

Several of the oncogenes activated during leukemia development affect regulators of gene expression. These include *C-MYC* and *TEL1* (genes that encode transcriptional activators) and *BCL6* (a gene that encodes a transcriptional repressor).

While most leukemias are sporadic and caused by translocations that are somatically acquired, a small proportion of leukemias arise in a manner that is clearly hereditary. Individuals affected by Fanconi anemia, ataxia aelangiectasia, Li Fraumeni syndrome and HNPCC are at significantly greater risk of leukemia, demonstrating that various forms of genetic instability can influence leukemia development.

Pancreatic Cancer

Cancer of the pancreas is the fifth leading cause of cancer death in the USA. It is a highly aggressive and lethal form of cancer, with a mortality rate that approaches the incidence. Known risk factors for pancreatic cancer include age, cigarette smoking and a family history of the disease.

The majority of cancers of the pancreas are adenocarcinomas that arise in the epithelia of the pancreatic ducts. The earliest lesions are known as *pancreatic intraepithelial neoplasia* (PanINs). PanINs been shown to be precursors of more advanced lesions in the epithelia of the pancreatic duct in a similar way that adenomas progress to invasive carcinomas in the colorectal epithelia. Because early pancreatic tumors tend to be associated with vague symptoms or are asymptomatic, most pancreatic cancers are advanced at the time of diagnosis. The pancreas is located deep in the abdominal cavity, in an anatomical region known as the retroperitoneal space (see Fig. 6.11). Symptoms of advanced cancers are related to obstruction of neighboring ducts and tracts (see Fig. 6.12).

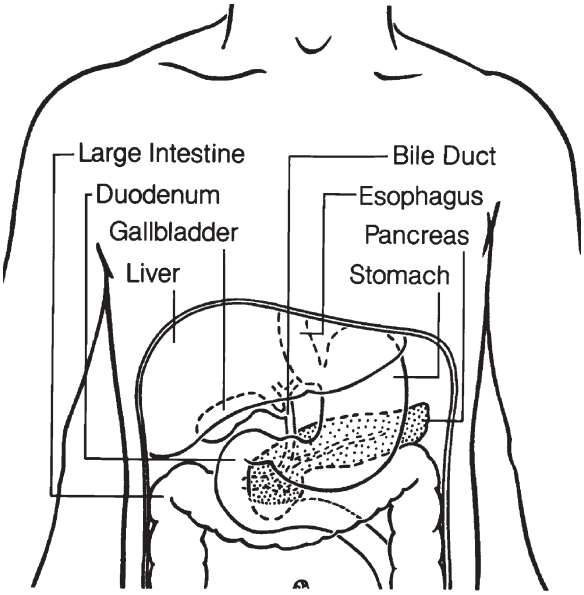

Fig. 6.11 Anatomy of the gastrointestinal tract. Cancer can arise at multiple sites in the gastrointestinal tract, including the pancreas, large intestine (colon), the gall bladder and bile duct, the liver and the stomach. (Courtesy of the National Cancer Institute.)

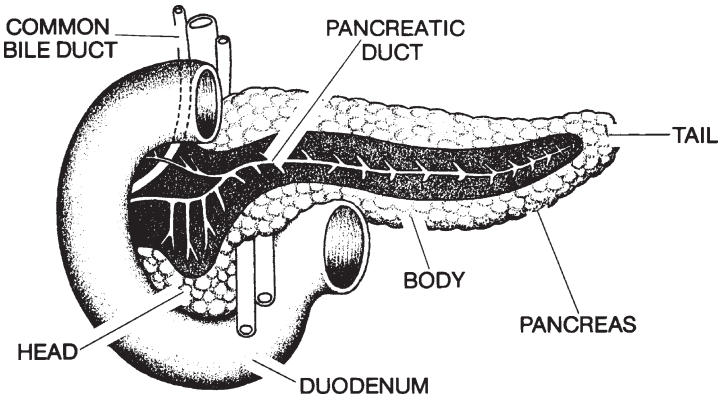

Fig. 6.12 Anatomy of the pancreas. About 60% of cancers arise in the head of the pancreas, 13% in the body and 5% in the tail. The remainder grow diffusely throughout the organ. Large carcinomas can constrict neighboring structures, such as the common bile duct and the duodenum. (Courtesy of the National Cancer Institute.)

Several genetic alterations have been found to be present in very high frequencies in pancreatic carcinomas. Activation of *K-RAS* by point mutation of codon 12 occurs in more than 90% of tumors. The small proportion of tumors that harbor wild type *K-RAS* alleles exhibits frequent activating mutations in *BRAF*. Among tumor

suppressor genes, the *CDKN2A* gene that encodes p16 is inactivated in nearly all pancreatic tumors, and *P53* and *SMAD4* are inactivated in more than half of all cases. Inactivation of *SMAD4* is specifically related to tumor suppression in the pancreatic epithelium, as it is mutated rarely in other malignancies. Less-frequent mutations have been reported in *RB, AKT2, TGFBR2* (the gene that encodes the TGF-β receptor) and in *MKK4* (a gene that encodes a kinase in the MAP kinase pathway).

As these patterns of mutation indicate, the RAS, TGF-β/ SMAD, p53 and the cell-cycle regulatory pathways are all altered in a large proportion of pancreatic cancers. In addition, mutations in the genome maintenance genes *BRCA2, FANCC* and *FANCG* have been found at lower frequencies.

Pancreatic cancer can arise in the setting of chronic inflammation (see Chapter 1). An inflammatory disease known as *hereditary pancreatitis* predisposes affected individuals to pancreatic cancer. Many cases of hereditary pancreatitis have been attributed to germline mutations in the *PRSS1* gene, which encodes a protease proenzyme called *cationic trypsinogen*. The most common mutations are of the missense type. Mutant proteins are inappropriately activated, leading to autodigestion of tissues in the pancreas, which triggers an inflammatory response (pancreatitis).

An estimated 10% of pancreatic cancers have an inherited component. In addition to hereditary pancreatitis, several known cancer syndromes predispose affected individuals to pancreatic cancer. A subset of individuals affected by the Familial Atypical Multiple Mole Melanoma (FAMMM) syndrome harbor germline mutations in *CDKN2A* and thereby has a significantly increased risk of pancreatic cancer. Pancreatic cancer is also part of the spectrum of cancers that occur as part of hereditary nonpolyposis colorectal cancer, Peutz–Jeghers syndrome, ataxia telangiectasia and the familial breast cancer syndrome caused by germline mutations in *BRCA2*.

Cancers of the Oral Cavity and Pharynx

Oral and pharyngeal cancers, including those of the lip, tongue, mouth and pharynx, are highly prevalent outside the USA, and rank as the sixth most common category of malignancy worldwide. In some countries, up to 40% of all cancers occur in the oral cavity and pharynx. In the USA, the major risk factors are smokeless tobacco and alcohol consumption.

The majority of oropharyngeal cancers are squamous cell carcinomas. These cancers typically arise in patients with preexisting dysplasia in the form of ulcers and plaques. Most early lesions are asymptomatic in their early stages. Developing tumors often grow laterally, sometimes becoming invasive and spreading to regional lymph nodes. Blood borne metastasis is uncommon.

P53 mutations are highly common in cancers throughout the head and neck, including the oropharynx. The hotspots at which the *P53* gene is mutated in oral and pharyngeal cancers overlap, but are partially distinct from, those found in the

lung and bladder cancers of cigarette smokers. Interestingly, the G→T transversions associated with benzo[a] pyrene exposure in lung cancers (see Chapter 1) are not commonly found in oropharyngeal cancers even though the mouth epithelia are exposed to the same agent.

Mutations in the RAS genes show an interesting pattern of bias among different ethnic and cultural groups. Rare among oral and pharyngeal cancers that occur in western countries and Japan, RAS gene mutations occur in a significant proportion of cancers that arise in India and Taiwan. Mutations in *H-RAS* are observed in Indian subjects who habitually chew tobacco, while *K-RAS* mutations are frequently detected in Taiwanese subjects who chew areca, a type of nut. The environmental or genetic factors responsible for such biases have not been determined.

Homozygous deletions that inactivate the p16-encoding gene *CDKN2A* are common in cancers of the oral cavity and pharynx. Also common is amplification of the region on chromosome 11q13 that contains the cyclin D (*CCND1*) locus.

While the cancers of the oropharynx are clearly associated with consumption of tobacco and alcohol, genetic factors that may be involved in disease susceptibility remain largely undefined. Familial clusters of oral cancers have been reported, but no germline cancer genes have been conclusively linked to risk of disease. Polymorphisms in genes that encode enzymes which participate in carcinogen detoxification, such as glutathione S-transferase (*GSTM1*), cytochrome 450 (*CYP1A1*) and aldehyde dehydrogenase (*ALDH2*) have been of interest, as several alleles have been found to be overrepresented in cancer patient cohorts.

Cancer of the Uterine Cervix

Carcinoma of the uterine cervix, commonly referred to as cervical cancer, ranks second only to breast cancer as a cause of cancer deaths worldwide. In the USA, the incidence has dropped considerably over the past several decades. This decline has been largely attributed to effective screening and treatment of precursor lesions. Among women with inadequate access to health care, particularly those in the developing world, cervical cancer remains highly prevalent.

The majority of cervical cancers are squamous carcinomas that arise in a region between two histologically distinct types of cervical epithelium, a region known as the *transition zone*. Disorganized lesions known as low-grade squamous intraepithelial *l*esions (LSILs) and high-grade squamous intraepithelial lesions (HSILs) are confined to the epithelia. HSILs are precursors that give rise to invasive squamous cell carcinomas.

In contrast to many of the common cancers in which cancer genes are sporadically acquired by somatic mutation or inherited, the most prevalent cancer genes in cervical cancers are acquired by infection. Cervical cancer is essentially an infectious disease. The causative agents are human papillomaviruses (HPV), small DNA-based viruses with a circular genome that is roughly 8 kb in size (see Fig. 6.13). Viral DNA sequences are found in over 90% of cervical cancers. Strains of HPV

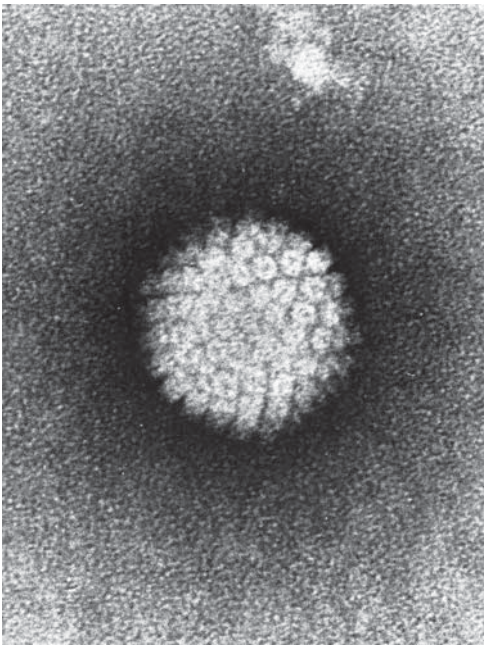

Fig. 6.13 Human Papilloma Virus (HPV). An electron micrograph of an HPV particle. HPV is a significant etiologic agent in the development of cervical cancer. (Courtesy of the National Cancer Institute.)

are distinguished by serotype. Relatively benign forms of disease, such as warts, are caused by the low-risk serotypes 6 and 11; highly malignant forms of this cancer are most strongly associated with the high-risk serotypes 16 and 18. Unlike more typical infectious diseases, only a small proportion of women who are infected with HPV develop cervical cancer.

Many of the viruses linked to cancer cause chronic inflammation. Cervical cancer is different in that HPV genes directly influence a cancer gene pathway. Two small viral open reading frames, E6 and E7, have been found to be strongly related to tumorigenesis. DNA containing E6 and E7 derived from high-risk serotypes can induce cancer-related phenotypes when transferred to cultured normal cells. Homologous DNA sequences from low-risk serotypes do not have this ability. Unlike the retroviral oncogenes described in Chapter 2, the E6 and E7 oncogenes of HPV have no known cellular homologs.

The E6 protein encoded by HPV binds strongly to p53. Upon binding, HPV E6 recruits a cellular protein, the E6-associated protein (E6-AP), a ubiquitin ligase. These three proteins form a complex that functions to ubiquitinate *p53* and thereby target it for degradation by the proteosome. Thus, HPV E6 facilitates the degradation of p53 and the consequent loss of p53 function in a fashion similar to the oncogene product MDM2 (see Chapter 3). In addition, the HPV E7 protein has

been shown to bind to and inactivate RB. The E6 and E7 proteins from high-risk HPV serotypes bind *p53* and RB more avidly than do the E6 and E7 proteins encoded by the low-risk serotypes. The simultaneous inactivation of two tumor suppressor genes by a cancer-causing virus strongly underscores the functional importance of these genes in maintaining tissue homeostasis within the epithelia of the cervix (see Fig. 6.14). The integrated virus benefits by the loss of growth control of its cellular host.

Other than HPV infection, the genetic alterations that drive cervical tumorigenesis remain largely unknown. Gains involving chromosome 3p24-29, resulting in the amplification of *PIK3CA*, are frequently observed in cervical neoplasia. A small proportion of cervical cancers harbor mutations in *K-RAS*. Other mutations in the PI3K/AKT or RAS pathways, in *PTEN* and *BRAF*, have not been reported. The *P53* gene is mutated in a small number of cervical carcinomas, indicating that in some cases the alteration of *P53* provides a additional selective advantage, even in the context of HPV infection.

Thyroid Cancer

Cancers of the thyroid are the most common malignancies of the endocrine system. Unlike many other cancers, rates of incidence of thyroid cancer have markedly increased over the past several decades. Known risk factors include exposure to ionizing radiation, reduced iodine intake, preexisting inflammatory disease and family history. The disease has a significantly higher incidence rate in females, possibly indicating a role for estrogen during tumorigenesis.

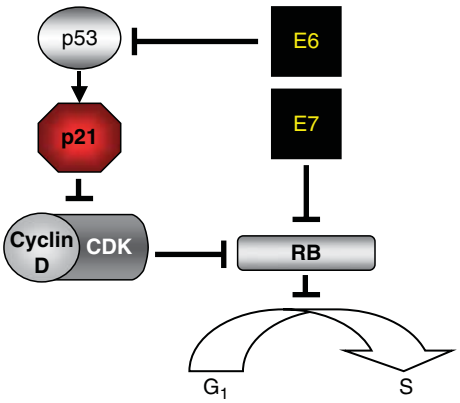

Fig. 6.14 Dysregulation of the $G_1 \rightarrow S$ transition during cervical tumorigenesis. HPV-encoded E6 and E7 proteins bind and functionally inhibit p53 and RB, respectively, leading to loss of $G_1 \rightarrow S$ checkpoint control in cells of the cervical epithelia

The majority of thyroid cancers arise from epithelial cells that line the thyroid follicles, which are responsible for synthesis of thyroid hormones. *Papillary carcinomas*, the most prevalent of the follicle-cell-derived neoplasia, arise as thyroid nodules (see Fig. 6.15). During tumorigenesis, thyroid tumors become progressively less developmentally differentiated and simultaneously more aggressive. Well-differentiated thyroid tumors can remain indolent for prolonged periods, and most of these do not progress. Undifferentiated thyroid tumors are among the most aggressive and lethal of all human cancers.

Fewer than 10% of thyroid cancers are *medullary carcinomas* that originate from a distinct cell type known as the *C-cell*. These tumors tend to grow slowly and have a favorable prognosis. Up to 25% of medullary thyroid cancers occur in individuals affected by the hereditary syndrome multiple endocrine neoplasia type 2.

Dysregulation of MAP kinase signaling (see Chapter 5) is a feature of nearly 70% of early, well-differentiated thyroid lesions. Loss of normal MAP kinase regulation most often results from nonoverlapping activating mutations in *RET* and *BRAF* and, less frequently, from *H-RAS* and *N-RAS* mutations. In contrast to the germline point mutations in *RET* that cause multiple endocrine neoplasia 2A (see Chapter 2), the activation of *RET* in most sporadic and inherited thyroid cancers occurs as a result of translocation.

Inactivating mutations of *P53* are relatively uncommon in thyroid cancers, occurring in 10% of tumors, most of which are poorly differentiated and aggressive. Mutations in *CTNNB1* that constitutively activate the WNT/APC pathway have similarly been implicated in the progression of thyroid tumors to invasive lesions.

Sporadic thyroid cancers attributed to radiation appear to have a distinct molecular etiology. A particularly high incidence of *RET* rearrangements has been observed in childhood thyroid cancers caused by radioactive fallout from the Chernobyl accident,

Fig. 6.15 Thyroid cancer nodules. Viewed as a gross specimen, the two lobes of a thryroid gland contain multiple cancerous nodules. (Courtesy of the National Cancer Institute.)

indicating a probable role for radiation-associated chromosome breaks. While the *BRAF* V600E point mutation occurs frequently in thyroid cancers, *BRAF* is altered primarily by translocation in thyroid cancers that are radiation related.

A positive family history increases risk threefold to sixfold indicating there is a significant heritable component to thyroid follicular-cell-derived carcinoma. Approximately 5% of all thyroid cancers can be attributed to a familial thyroid cancer syndrome, the genetic etiology of which remains unknown. Thyroid cancers are also a part of the clinical spectrum of familial adenomatous polyposis, Cowden syndrome, and Werner syndrome. However, mutations in *APC, PTEN* and *WRN* are not found in sporadic tumors.

Stomach Cancer

Stomach cancer was once the leading cause of cancers deaths in the USA, but its incidence in developed countries has decreased significantly over the past several decades. This disease remains a major cause of cancer death in many developing countries, and ranks just below lung cancer as the second most common cause of total cancer deaths worldwide. Environmental factors play a significant role in stomach cancer risk.

Stomach cancers are strongly associated with chronic inflammation (see Chapter 1). An important etiologic agent is the bacterium *Helicobacter pylori*, which causes ulcerative disease. Stomach cancer is particularly prevalent in developing countries with a high incidence of *H. pylori* infection early in life. The inflammatory lesion *chronic atrophic gastritis* is a precursor to invasive cancers. Other disease states that involve both inflammation of the gastric mucosae and increased cancer risk are pernicious anemia and Epstein Barr Virus infection.

By far the most predominant cancer of the stomach is *gastric adenocarcinoma*, which arises from the glands of the gastric epithelium. Two histologically distinct forms of gastric adenocarcinoma have been identified: the intestinal type that forms gland-like structures and a diffuse form that is infiltrative in nature. The intestinal form arises most typically in older individuals from precursor lesions, and spreads via bloodstream to the liver. In contrast, the diffuse form of gastric adenocarcinoma tends occur in all age groups, has no identifiable precursor lesions and spreads mainly into contiguous tissues. Most gastric adenocarcinomas present as advanced lesions, and therefore have a high mortality rate. The intestinal form is more closely associated with environmental agents, and cancers of this type appear to acquire a distinct pattern of genetic alterations. The diffuse form has been more closely linked to heritable factors, as will be described below.

Several recurrent chromosomal regions of loss of heterozygosity occur in stomach cancer, but in most cases, these regions of loss have not been definitively linked to known tumor suppressor genes. Mutations of *P53* have been detected in approximately 30% of gastric carcinomas, sometimes in precursor lesions but more frequently in advanced cancers. A small number of *APC* mutations have been

reported, particularly in Japanese patients. Alteration of the transforming growth factor-β (TGF–β) type II receptor occurs frequently, particularly in the large subset of stomach cancers that are mismatch repair deficient.

Approximately 10% of stomach cancers occur in familial clusters, and 1–3% are clearly hereditary in nature. Notably, the family of Napoleon Bonaparte was afflicted with a dominantly inherited form of stomach cancer. The best-defined familial stomach cancer syndrome is *hereditary diffuse gastric cancer* (HDGC). About 30% families with HDGC carry germline mutations in *CDH1*, the gene that encodes E-cadherin, an important mediator of cell–cell adhesion (see Chapter 5). Mutations throughout the *CDH1* open reading frame confer predisposition to the diffuse form of gastric adenocarcinoma. The predominant type of *CDH1* mutations produce premature termination codons. The penetrance of mutant *CDH1* alleles is high; in the Maori kindred in whom the disease was first identified, more than 25 individuals have died of stomach cancer, the youngest of whom was age 14. Affected individuals are also at significantly increased risk of lobular breast cancer. *CDH1* is a good example of a potent tumor suppressor gene that is involved in predominately one type of cancer.

The most prevalent genes that increase the risk of stomach cancers are the mismatch repair genes. Many stomach cancers exhibit evidence of microsatellite instability and stomach cancer is a significant component of the hereditary nonpolyposis colorectal cancer disease spectrum.

Brain Tumors

In the USA, approximately 20,000 individuals are diagnosed with primary brain cancer each year. Brain cancers occur in all age groups; over 8% of brain tumor deaths occur in individuals younger than 34 years.

The majority of malignant tumors that occur within the central nervous system are not derived from neurons, but arise from supportive tissues known as the *neuroglia*. The neuroglia are composed of glial cells, several types of which can develop into tumors. The most common type of glial cell tumor is the *astrocytoma*. Astrocytomas can occur in all areas of the brain and spinal cord of children and adults (see Fig. 6.16), and can be classified in four distinct grades. The most malignant form of astrocytoma is the grade IV tumor, also known as the *glioblastoma multiforme* (GBM). GBM are the most common primary brain cancers in adults, and among the most lethal cancers of any type (see Fig. 6.17).

As the name multiforme implies, GBMs are morphologically heterogenous. GBM appears to arise via two distinct pathways. Some GBMs rapidly arise in older individuals, often in the cerebral hemispheres, in the absence of any precursor lesion. These tumors are known as primary, or *de novo* GBMs. In contrast, secondary, or progressive, GBMs arise as lower-grade astrocytomas and slowly progress to more aggressive cancers. The two categories of GBM harbor distinct but overlapping patterns of genetic alterations. Primary GBMs frequently harbor deletions of *PTEN*

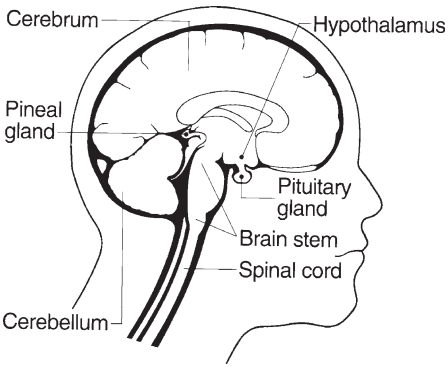

Fig. 6.16 Anatomy of the central nervous system. Cancer can arise in various tissues within the brain and spinal cord. (Courtesy of the National Cancer Institute.)

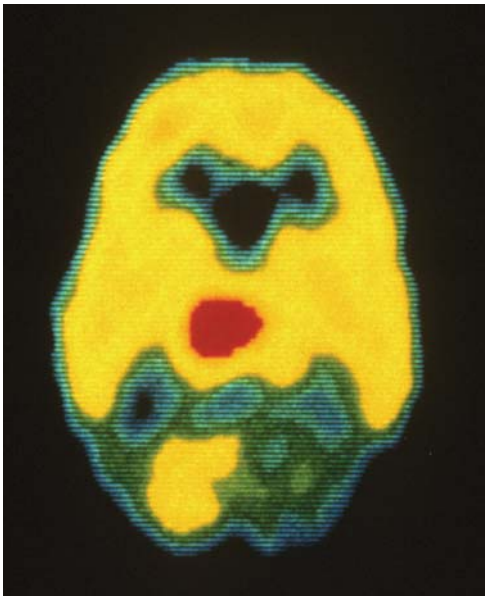

Fig. 6.17 A malignant brain tumor. Tumors in the brain that are highly aggressive use more glucose than surrounding normal tissues. An area of high glucose uptake within a tumor (red) can be visualized by positron emission tomography (PET). (Courtesy of the National Cancer Institute.)

and *CDKN2A*, and amplification of *EGFR*. Secondary GBMs most commonly exhibit inactivating *P53* mutations, particularly those that alter codon 273. Activating point mutations in *PIK3CA* have been found in approximately 15% of both types of GBM.

As with other common cancers, the majority of GBMs are sporadic and arise as a result of somatically acquired mutations. Inherited GBMs occur as a component

of familial adenomatous polyposis, Li–Fraumeni syndrome, and central neurofibromatosis (see Chapter 3).

Medulloblastomas, the most common primary malignant brain tumor among children, arise in the cerebellum. Approximately 5% of medulloblastomas exhibit cytogenetic evidence of gene amplification. In many such cases, the *C-MYC* or *N-MYC* loci are included in the amplicon. Medulloblastomas are associated with familial adenomatous polyposis, particularly of the subtype known as Turcot syndrome (see Chapter 3). While *APC* gene mutations have not been reported in sporadic medulloblastomas, the WNT/APC pathway is dysregulated in a small proportion of sporadic tumors by activating *CTNNB1* mutations.

The second most common adult brain cancer is the *menigioma*, derived from glial cells in the meningeal coverings of the brain and spinal cord. Menigiomas are a component of the neurofibromatosis type 2 cancer syndrome (see Chapter 3); greater than one half of sporadic meningiomas exhibit somatically acquired mutations in the *NF2* gene.

Less common glial cell-derived brain cancers include *oligodendrogliomas* and *ependymomas* which arise in the the cerebral hemispheres, and the linings of the ventricles, respectively. Few consistent genetic alterations have been found in these uncommon forms of brain cancer.

Liver Cancer

While relatively uncommon in the USA, liver cancer is one of the most common and most deadly malignancies worldwide. The highest incidence of this disease is found in southern China and sub-Saharan Africa, where it accounts for as many as 10% of deaths from all causes. The major risk factors are chronic infection by the hepatitis B virus (HBV) and hepatitis C virus (HCV), and food contamination with aflatoxin B1 (see Chapter 1). Liver cancers typically arise in the predominant epithelial cells in the liver, the hepatocytes, and are therefore referred to *hepatocellular carcinomas* (HCC).

The majority of HCCs feature the functional inactivation of p53-dependent transcription. There are several mechanisms by which loss of p53 function occurs. Exposure to various chemical and environmental toxins causes characteristic inactivating mutations in the *P53* gene. The best known example of toxin-specific mutagenesis is caused by the dietary toxin aflatoxin B1, which frequently contaminates corn, rice and peanuts. Aflatoxin B1 causes G→T transversions at *P53* codon 249, a site rarely mutated in other cancer types (see Chapter 1). In areas of low aflatoxin B1 intake, a distinct spectrum of *P53* mutations is observed. In HBV-associated HCC, p53 protein interacts with the product of a viral oncogene known as gene X, or *HBX. HBX* is expressed as a result of the frequent integration of the HBV genome into the DNA of the host cells. The *HBX*-encoded protein directly binds p53 and inhibits the association of *p53* with its target sequences in the genome, thus repressing p53-mediated transcription. Besides the mutations that

affect *P53*, the genetic alterations that contribute to tumorigenesis in the liver remain poorly understood.

HBV appears to promote HCC in two different ways: (1) the HBV *HBX* gene directly inhibits the p53 pathway and (2) HBV infection causes chronic inflammation that appears to be a general mediator of malignant transformation. In contrast to HBV, the genome of HCV does not encode a functional homolog of *HBX*. It therefore appears that the risk of HCC caused by HCV infection is entirely attributable to the effects of chronic inflammation. Other HCC risk factors that trigger chronic inflammation are alcohol-induced liver disease, and the genetic diseases hemochromatosis (an iron overload disease), Wilson disease (a copper overload disease), porphyria (a heme-pigment overload disease) and α1-antitrypsin deficiency.

Further Reading

Bell, D. A. Origins and molecular pathology of ovarian cancer. *Mod. Pathol*. **18** (Suppl 2), S19–S32 (2005).

de Snoo, F. A. & Hayward, N. K. Cutaneous melanoma susceptibility and progression genes. *Cancer Lett*. **230**, 153–186 (2005).

Ellenson, L. H. & Wu, T. C. Focus on endometrial and cervical cancer. *Cancer Cell* **5**, 533–538 (2004).

El-Rifai, W. & Powell, S. M. Molecular biology of gastric cancer. *Semin. Radiat. Oncol*. **12**, 128–140 (2002).

Farazi, P. A. & DePinho, R. A. Hepatocellular carcinoma pathogenesis: From genes to environment. *Nat. Rev. Cancer* **6**, 674–687 (2006).

Gallia, G. L. et al. PIK3CA gene mutations in pediatric and adult glioblastoma multiforme. *Mol. Cancer Res*. **4**, 709–714 (2006).

Haluska, F. G. et al. Genetic alterations in signaling pathways in melanoma. *Clin. Cancer Res*. **12**, 2301s–2307s (2006).

Hovey, R. M. et al. Genetic alterations in primary bladder cancers and their metastases. *Cancer Res*. **58**, 3555–3560 (1998).

Hussain, S. P., Schwank, J., Staib, F., Wang, X. W. & Harris, C. C. TP53 mutations and hepatocellular carcinoma: Insights into the etiology and pathogenesis of liver cancer. *Oncogene* **26**, 2166–2176 (2007).

Kangelaris, K. N. & Gruber, S. B. Clinical implications of founder and recurrent CDH1 mutations in hereditary diffuse gastric cancer. *JAMA* **297**, 2410–2411 (2007).

Kondo, T., Ezzat, S. & Asa, S. L. Pathogenetic mechanisms in thyroid follicular-cell neoplasia. *Nat. Rev. Cancer* **6**, 292–306 (2006).

Linehan, W. M., Walther, M. M. & Zbar, B. The genetic basis of cancer of the kidney. *J. Urol*. **170**, 2163–2172 (2003).

Maitra, A., Kern, S. E. & Hruban, R. H. Molecular pathogenesis of pancreatic cancer. *Best Pract. Res. Clin. Gastroenterol*. **20**, 211–226 (2006).

Moasser, M. M. The oncogene HER2: Its signaling and transforming functions and its role in human cancer pathogenesis. *Oncogene* (2007).

Munger, K. et al. Mechanisms of human papillomavirus-induced oncogenesis. *J. Virol*. **78**, 11451–11460 (2004).

Nakayama, K. et al. Amplicon profiles in ovarian serous carcinomas. *Int. J. Cancer* **120**, 2613–2617 (2007).

Nathanson, K. L., Wooster, R. & Weber, B. L. Breast cancer genetics: What we know and what we need. *Nat. Med*. **7**, 552–556 (2001).

Nylander, K., Dabelsteen, E. & Hall, P. A. The p53 molecule and its prognostic role in squamous cell carcinomas of the head and neck. *J. Oral Pathol. Med*. **29**, 413–425 (2000).

Oda, K., Stokoe, D., Taketani, Y. & McCormick, F. High frequency of coexistent mutations of PIK3CA and PTEN genes in endometrial carcinoma. *Cancer Res*. **65**, 10669–10673 (2005).

Ohgaki, H. & Kleihues, P. Genetic pathways to primary and secondary glioblastoma. *Am. J. Pathol*. **170**, 1445–1453 (2007).

Sagaert, X., De Wolf-Peeters, C., Noels, H. & Baens, M. The pathogenesis of MALT lymphomas: Where do we stand? *Leukemia* **21**, 389–396 (2007).

Warnakulasuriya, K. A. & Ralhan, R. Clinical, pathological, cellular and molecular lesions caused by oral smokeless tobacco – a review. *J. Oral Pathol. Med*. **36**, 63–77 (2007).

Chapter 7
Cancer Genetics in the Clinic

The Uses of Genetic Information

The cancer gene theory has provided an intellectual framework for understanding how cancers arise and how they grow. That mutated genes provide selective advantages at various stages of tumor growth explains how tumorigenesis is related to both our environment and our inborn genetic makeup. These insights rank among the great accomplishments of modern science. Most importantly, the cancer gene theory guides the most promising efforts to prevent, diagnose, treat and cure cancer.

The clinical uses of genetic information are many and varied. This final chapter will briefly highlight several of the practical applications of cancer genetics:

Genetic testing. Inherited cancer genes can significantly increase the lifetime risk of developing cancer (see Chapter 3). Therefore, the identification of well-characterized germline cancer genes can be used to predict both the type and extent of cancer susceptibility.

Diagnosis and prognosis. Many cancers have a course that is highly variable and therefore difficult to predict. Because cancer genes dictate the aberrant phenotype of cancer cells, genotypic analysis can provide information on the capacity of a given tumor to grow and spread. This information can potentially be used to categorize tumors and to predict their course and responses to therapy.

Early detection. The most treatable cancers are those that are diagnosed at an early stage. Molecular genetic methods have the potential to detect cancer cells with high sensitivity and high specificity.

Rational therapies. Cancer genes are the essential difference between cancer cells and their normal neighbors. Insights into cancer genes and the pathways they control provide molecular targets, against which highly specific therapies can be designed. The term ‘magic bullet’, originally coined in the 1800s by the bacteriologist Paul Ehrlich to describe a drug that would specifically target pathogenic microorganisms, has since been applied to drugs that can target cancer cells harboring specific cancer genes.

F. Bunz, *Principles of Cancer Genetics*.
© Springer 2008

Elements of Cancer Risk: Carcinogens and Genes

Cancers are not evenly distributed throughout the entire human population. For many cancers, the rate of incidence is clearly higher in some identifiable groups of people than in others (see Chapter 6). These differences in risk are attributable to both variable exposure to carcinogens and genetic inheritance and, in some cases, to synergistic combinations between these two factors. The development of sun-related skin cancers in individuals affected by xeroderma pigmentosum exemplifies the interplay between genes and the environment (see Chapter 4).

Carcinogen exposure can be linked to multiple aspects of lifestyle and geographic location. Lung cancer occurs disproportionately in individuals who smoke; liver cancer has a high incidence among groups of individuals who live in areas where aflatoxin B1 and the hepatitis B virus are endemic. There are many more examples in which the prevalence of cancer is linearly related to exposure to a known carcinogen. In addition, there are numerous environmental factors that contribute to cancer risk in subtle ways that remain incompletely understood.

While nongenetic factors are clearly important, genetic makeup is a significant component of cancer risk. This conclusion is based upon a large amount of molecular and epidemiological evidence. While the majority of human cancers are sporadic, many cancers cluster in families. Some familial clusters occur by chance, most often in cases in which the type of cancer is a common one. When clusters are nonrandom, familial cases may reflect underlying germline cancer genes.

An estimated 5–10% of all cancers can be attributed to a heritable mutation in a high-penetrance cancer gene. In most cases, such genes are tumor suppressor genes (see Chapter 3). Individuals that inherit mutated, inactivated tumor suppressor alleles tend to develop benign and/or malignant tumors at a young age, and in many cases develop more than one primary tumor.

Alternatively, a familial cluster of cancer might result from the cumulative effect of multiple low-penetrance cancer genes. Because low-penetrance cancer genes are inherently more difficult to identify and to assess, fewer of these genes have been characterized extensively. The overall impact of low-penetrance cancer genes has yet to be determined, but clearly has the potential to be significant.

Identifying Carriers of Germline Cancer Genes

Many known cancer genes confer quantifiable cancer risks; the list of such alleles is rapidly growing. The tools of molecular biology have served to identify risk factors and these tools also present the means to evaluate risk. There are two central questions: What alleles are practically informative? Who should be tested? One way to determine whether to test a given individual or group for the presence of a

defined cancer gene is to weigh the benefits that genetic information might provide against the cost of obtaining that information.

Benefits of carrier identification. The identification of carriers of germline cancer genes can be of great benefit, both to the carriers themselves and to their families. Carriers of well-characterized cancer genes are at significant risk for specific types of cancer. In such cases, a positive genetic test constitutes an early warning and indicates close monitoring of that patient and further testing of that patient's close relatives. Depending on the gene, some carriers may well have latent tumors that warrant immediate therapy. There is a higher probability that treatment will be successful if delivered before the onset of symptoms.

Chemopreventive drugs can partially ameliorate the risk of some types of cancer. In other cases, prophylactic surgery is an option. For example, the diagnosis of familial adenomatous polyposis carries a near-certain risk of colorectal cancer. Family members who carry high-penetrance *APC* mutations often elect to undergo total colectomy, greatly reducing their risk of cancer. Similarly, the high risk of breast cancer conferred by germline mutations in *BRCA1* and *BRCA2* can be lowered by prophylactic mastectomy.

At the other extreme, there are diseases in which definitive genetic information has little impact on the medical management of carriers. A classic example is Huntington disease, a degenerative neurological disorder caused by the autosomal dominant inheritance of mutations in the gene *HTT*. *HTT* mutations have very high penetrance and therefore provide a powerful estimate of disease risk. The problem with *HTT* mutation testing is that despite advances in analytical technology, there remains nothing that can be done to prevent Huntington disease or to alter its course. Whether or not to test becomes a decision highly based on personal preference. As aptly stated by the blind seer Tiresias in *Oedipus the King*, by Sophocles, 'It is but sorrow to be wise when wisdom profits not'. The ability to test for a genetic condition for which there is no intervention gives rise to a unique dilemma.

For almost every type of cancer, some form of intervention is possible at the early stages of disease. Therefore, there is a generally a benefit to identifying at-risk individuals. The magnitude of that benefit is dependent on the type of cancer.

Costs of carrier identification. The allele frequencies of cancer genes are fortunately low. Even the more prevalent cancer genes, such as mutant alleles of *ATM*, are present in fewer than 1% of individuals in the general population. The low frequency of known cancer genes in the general population necessitates the testing of many non-carriers for every carrier who is ultimately identified.

Most modern tests for cancer genes involve the direct sequencing of genomic DNA. The conclusive identification and confirmation of mutations typically requires multiple DNA amplification and DNA sequencing reactions that must be performed by highly trained personnel using specialized laboratory equipment. At the present time, this type of genetic testing is expensive and therefore restricted to individuals in developed countries.

The low yield and high expense of cancer gene screening combine to present a significant barrier to the widespread application of genetic testing to the

general population. Even in a wealthy nation such as the USA, the monetary resources dedicated to health care are finite. Money spent on large-scale screening efforts would ultimately divert funds away from direct patient care, and from programs designed to improve cancer prevention and education and further cancer research. Large-scale genetic testing is only feasible if it *lowers* the overall cost of health care.

Improvements in technology promise to have a significant impact on the cost of genetic testing and population screening. Advances in genomic DNA sequencing and DNA sequence analysis tend to decrease the unit cost of genetic information. The simple economies of scale also promise to significantly lower DNA sequencing costs. As gene analysis efforts grow in size and DNA sequencing technology continues to evolve, broader population-based screens will become increasingly economically feasible.

Balancing benefits and costs. The benefit of large-scale cancer gene screens to society is dependent on their cost-effectiveness. Screening efforts focused on at-risk populations, in which cancer gene allele frequencies are relatively high, can significantly increase the yield of carriers identified. Thus, the cost per carrier identified is lower when screens are applied to at-risk groups. Furthermore, founder effects within such groups can result in the frequent appearance of highly characteristic alleles (with mutations is predictable regions of a susceptibility gene), minimizing the amount of DNA that must be sequenced.

The cost-benefit considerations of genetic testing are well illustrated by the case of familial breast cancer. The discovery of *BRCA1* and *BRCA2* in the 1990s made it possible to test for breast cancer susceptibility. Collectively, mutations in the BRCA genes occur in the general population at a frequency of about 1 in 250 women. However, defined subpopulations harbor mutant BRCA genes at a much higher frequency. A high proportion (10–12%) of *BRCA1* or *BRCA2* mutations occur in families with two or more cases of early onset breast cancer, women with invasive ovarian cancer and Jewish women with breast cancer.

Among Jewish women, founder effects have resulted in a high frequency of three characteristic breast cancer mutations, two in *BRCA1* (185delC and 5382insC) and one in *BRCA2* (6174delT). If a Jewish woman does not carry one of these founder mutations, it is improbable that a different BRCA mutation will be detected. Therefore, sequencing just three known regions in *BRCA1* and *BRCA2* would detect the >10% of at-risk individuals in this population. Other founder mutations have been identified in Icelandic and Polish subpopulations. The presence of characteristic, high-penetrance alleles in defined populations greatly facilitates efforts to identify carriers in a cost-effective manner.

Altered Genes as Biomarkers of Cancer

A defined genetic alteration that can be used to assess the risk or presence of disease is an example of an assayable cellular feature known as a *biomarker*. Biomarkers can be used to determine the risk of cancer, to screen for cancer and

confirm the presence of suspected cancer, and to determine the prognosis or staging of cancer. Biomarkers can also be used to monitor and optimize treatment by providing oncologists information they can use to avoid futile therapy, or to dose specific treatments with more precision. Both inherited and somatically acquired cancer genes can serve as biomarkers. The analysis of inherited alleles can provide critical information regarding cancer risk, as described in the previous section. The detection of somatically acquired cancer genes can potentially provide an additional parameter to evaluate an existing neoplasm, with which an oncologist can make a definitive diagnosis and establish a prognosis.

The detection of a tumor immediately raises several critical questions. What type of cancer is this? What is the most probable future course of this cancer? How will this cancer respond to therapy? Oncologists assess many different parameters of a newly discovered tumor, including size, location and spread, cellular composition and cellular appearance, as they attempt to predict the future course of the disease. In general, tumors are compared with previously documented tumors that appear to be similar. Following detailed analysis and evaluation by highly experienced physicians, uncertainties often remain. Tumors that appear similar can subsequently exhibit very different clinical courses that lead to different outcomes. In many instances, there is simply not enough information available to distinguish one tumor from another.

The use of cancer genes as biomarkers for cancer has many current and potential applications:

Diagnosis. The genetic etiology of a cancer can define a cancer type. The classic example is chronic myelogenous leukemia (CML), in which the observation of the Philadelphia chromosome is diagnostic (see Chapter 2). The polymerase chain reaction (PCR), a method used to amplify short DNA sequences, is used to detect the hybrid *BCR-ABL* oncogene that is present in 95% of CML patients. The proportion of Philadelphia chromosome-positive cells present in the blood and bone marrow is directly proportional to the total expression of *BCR-ABL*. Therefore, the response of CML to therapy can be monitored by assessing the levels of *BCR-ABL* genomic copies by standard PCR or *BCR-ABL* RNA transcripts by quantitative reverse transcription-PCR.

Staging. The extent to which a tumor has spread to distant tissues, or *metastasized*, is a portentous prognostic sign that has significant implications for patient management. Staging is particularly important in evaluating tumors of the breast, as the development of metastases is the predominant cause of death from breast cancer. The detection of disseminated cancer cells at or around the time of surgery is an indication for more aggressive adjuvant therapy, including chemotherapy and radiation.

At the time of surgery, lymph nodes that drain the affected breast (known as *sentinel nodes*) are dissected for evidence of metastatic disease. Methods that rely on microscopy to detect disseminated cancer cells in nodal tissue are relatively insensitive; up to 30% of patients judged to be node-negative by traditional methods develop distant metastases within 5 years. PCR-based methods have the potential to detect small numbers of disseminated cancer cells, referred to as *micometastases*, with high sensitivity and high specificity. Such assays are

designed to detect genes that are overexpressed and/or altered in breast cancer cells, including *ERBB2* and *EGFR*.

Recurrence. A challenge often faced during the course of treatment of cancers is the recurrence of disease after surgery. Cancers of the oral cavity and pharynx exhibit a high recurrence rate after surgical excision. Recurrence can be attributed to a small number of cancer cells that remain on the margins of the excised region. The squamous cell cancers that are most common in the oropharynx, for example, frequently contain *P53* mutations. In such cases, the detection of frequently observed mutated *P53* alleles would have the potential to improve detection of residual cancer cells.

Prognosis. Can the spectrum of genetic alterations in a tumor presage the future of a cancer patient? Genetic information has the potential to significantly factor into disease prognosis. An illustrative example is the *P53* gene, which is mutated in a high percentage of many types of tumors. Because *P53* mutations are so prevalent in cancers, the have been numerous attempts to establish the extent to which somatic *P53* mutations are predictive of disease progression and response to therapy. These studies have used different forms of technology and yielded results that have often been difficult to interpret. Nonetheless, there is an emerging consensus that somatic *P53* mutations correlate with progression to an advanced cancer and portend an unfavorable outcome in several common cancers.

A good example of a disease that is progressively linked to *P53* loss is Barrett's esophagus. Barrett's esophagus is an established precursor to esophageal adenocarcinoma. Whereas most patients with Barrett's esophagus do not progress to cancer, patients that do progress have a poor prognosis. Numerous studies have explored the use of *P53* status to predict the progression of a noninvasive lesion to invasive cancer. Current management entails periodic endoscopic examination and tissue biopsies. Given the relatively high prevalence of Barrett's esophagus but low overall risk of progression, the invasive and expensive approach currently in use is not cost-effective. A genetic approach would be highly applicable to this problem. Mutated *P53* is frequently observed in esophageal adenocarcinomas, but is uncommon in earlier precursor lesions. As in colorectal cancer, it appears that loss of *P53* occurs when esophageal neoplasia begin to invade surrounding tissues. A lack of a detectable *P53* mutation cannot rule out progression of Barrett's esophagus. However, mutant *P53* may be useful as an early marker to identify the individuals in whom Barrett's esophagus is most likely to progress. Such individuals would greatly benefit from close surveillance.

Another disease that has been highly studied in this regard is breast cancer. In numerous studies, *P53* mutations have been shown to predict an unfavorable prognosis. The predictive value of a *P53* mutation appears to be independent of other prognostic factors such as tumor size, lymph node status and expression of the estrogen receptor. Mutations that alter the DNA binding domain of p53, and thus effect transcriptional transactivation, appear to be associated with worse prognosis that those that occur outside this domain.

Multiple biomarkers can be simultaneously assessed in order to increase the amount of information obtained from a molecular assay. For example, *P53* mutation

status has been shown to be useful in identifying women at higher risk of disease recurrence and death when their tumor also had amplification of *ERBB2*, which is an independent prognostic marker. In principle, the parallel assessment of a sufficiently large number of informative molecular markers could provide both the clinician and the cancer patient a detailed view into the future.

Detecting Early Cancers via Gene-Based Assays

Cancers that are detected at early stages of tumorigenesis are most likely to respond to curative therapies. For types of tumors that grow in a stepwise manner, early lesions have not yet acquired all of the genetic alterations that give rise to aggressive, metastatic growth. Such lesions are ideally suited for surgical resection and tend to be sensitive to other forms of therapy. For example, smaller, noninvasive colorectal tumors are less likely than larger tumors to carry *P53* mutations (see Chapter 1). Loss of *P53* is associated with therapeutic resistance, and thereby contributes to an unfavorable outcome. The detection of early, *P53*-proficient tumors is therefore an important goal.

A highly sensitive means of detecting early tumors would have a significant impact on cancer mortality. However, depending on the type of cancer, the potential for the diagnosis of early, noninvasive lesions is highly variable. Tumors of the skin, such as melanomas, can be detected visually and are thus often diagnosed at early stages. Noninvasive tests such as mammograms effective screens for the detection of smaller tumors of the breast, while chest radiographs have been a somewhat less-reliable means of detecting small tumors in the lung. Tumors in the colorectum and upper gastrointestinal tract can be detected by more invasive endoscopic procedures. Highly lethal tumors such as pancreatic and ovarian cancers are not often detected on routine examination and are typically diagnosed at advanced stages of disease.

Because all cancer cells carry cancer genes, one attractive approach to the diagnosis of early-stage tumors is the detection of specific mutations in clinical samples. Cancer cells are continuously sloughed from the surfaces of growing tumors into various bodily fluids and tissue spaces. In many cases, the genetic content of these cells can be analyzed by techniques involving PCR. By selectively querying genes known to be mutated at high frequency in a given type of cancer, a highly sensitive and specific diagnosis is possible. A limiting factor is the preponderance of genetically normal cells that are invariably present in clinical samples, which can obscure the presence of cancer cells. Several techniques have been developed to detect cancer genes against a background of more numerous normal genes. This general approach has been applied experimentally to several common types of cancer:

Lung cancer. At the time of diagnosis, more than 65% of all patients with non-small-cell lung cancer will have advanced disease that is no longer amenable to curative therapy. Early diagnosis would identify patients with potentially resectable

disease. Molecular screening for lung cancer has focused on detecting exfoliated cells in several bodily fluids, including sputum and the fluid obtained during bronchoalveolar lavage. Cancer cell DNA can also be detected in serum and plasma samples from the circulatory system.

The gene most commonly queried in lung cancers has been *K-RAS*. *K-RAS* is mutated in nearly 50% of all primary lung adenocarcinomas and in the majority of cases mutations affect a single residue, encoded by codon 12 (see Chapter 6). Clinical materials obtained from the lung contain many cells, a large proportion of which are inflammatory cells with normal genes. Several strategies have been employed to enrich for mutant *K-RAS* genes in complex solutions of normal DNA (see Fig. 7.1).

Colorectal cancer. While colorectal cancer is among the leading causes of cancer death in the US, noninvasive tumors are highly curable. Screening for early colorectal tumors is therefore critical to reducing the overall impact of this disease. A widely used screen that tests for fecal occult blood is noninvasive but suffers from both low sensitivity and specificity. Among the invasive tests available are

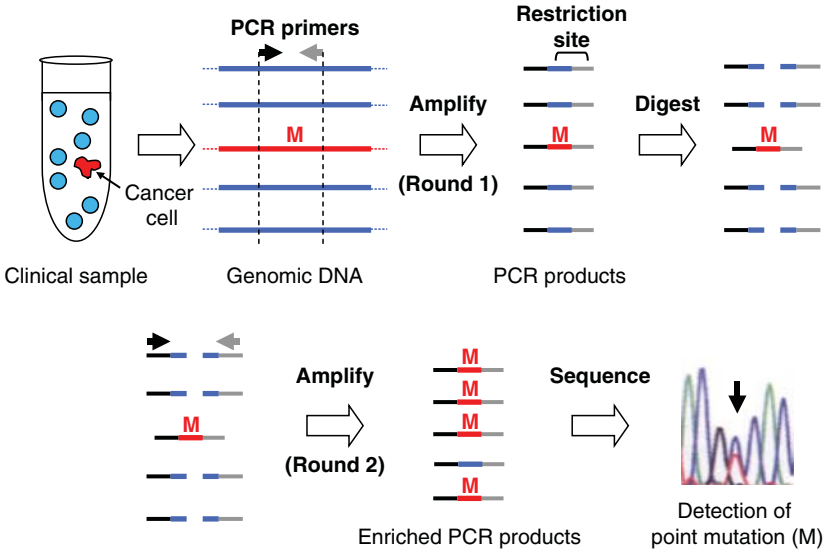

Fig. 7.1 Sensitive detection of rare *K-RAS*-mutant cancer cells. Several strategies have been employed to allow the detection of rare cancer cells. In this example of such an approach, a clinical sample such as fluid from a bronchoalveolar lavage contains a cancer cell (red) as well as many cells that are genetically normal (blue). Lung cancer cells frequently harbor mutations in *K-RAS* in codon 12 (red allele marked 'M'). DNA primers (arrows), designed to amplify this small region, also contain DNA sequences that complete a recognition site for a DNA endonuclease (restriction enzyme). Because, this engineered recognition site is absent in PCR products that carry the codon 12 mutation, incubation with the specific endonuclease results in the preferential digestion of the wild type-derived PCR products. Cut DNA is not efficiently amplified during a second round of amplification. A population of amplified DNAs enriched for the mutant DNA can be detected by DNA sequencing

colonoscopy and barium enemas followed by radiography. While highly sensitive and specific, these methods are expensive and uncomfortable, limiting patient compliance.

The ideal molecular marker in colorectal cancer is the mutant *APC* gene, which is present in the large majority of tumors, at all stages of growth. *K-RAS* mutations, in contrast, are present in most growing colorectal tumors but are also found in benign neoplasia that are at low risk of progression (see Chapter 2). Unlike *K-RAS* mutations that must often occur at a single codon, *APC* mutations occur throughout the first 1,600 codons of the gene (see Chapter 3), and therefore cannot be reliably detected with a single generic sequencing reaction. Because the majority of epithelial cells that are sloughed into the bowel lumen are genetically normal, mutant *APC* alleles account for fewer than 1% of the total recovered from fecal samples. Both of these obstacles are circumvented by an experimental diagnostic assay called the digital protein-truncation test (see Fig. 7.2). This assay reduces the complexity of fecal DNA by dividing it into smaller pools, thereby allowing the detection of

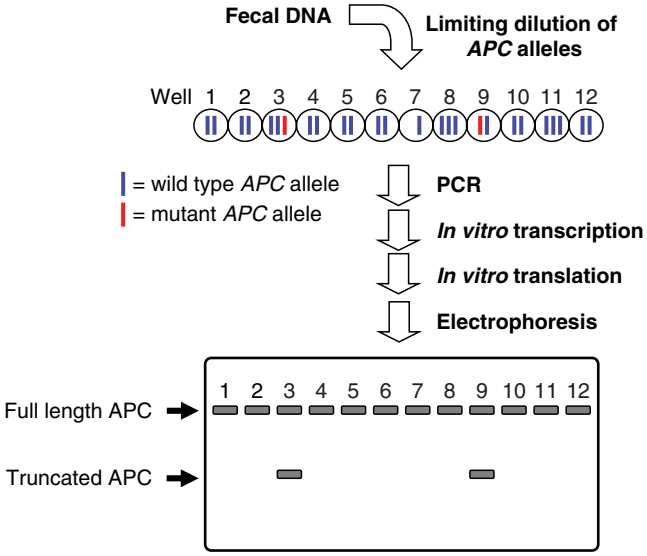
Fecal DNA
Limiting dilution of APC alleles
Well 1 2 3 4 5 6 7 8 9 10 11 12
= wild type APC allele
= mutant APC allele
PCR
In vitro transcription
In vitro translation
Electrophoresis
1 2 3 4 5 6 7 8 9 10 11 12
Full length APC
Truncated APC

Fig. 7.2 Detecting tumor-associated mutant APC genes in a stool sample. The Digital Protein-Truncation test allows diverse, mutant *APC* alleles (red) to be detected against a background of more numerous wild type *APC* alleles (blue) derived from normal cells. To reduce the complexity of the PCR template and obtain a detectable signal, fecal DNA is distributed into multiwell plates at limiting dilution, so that an average well contains two or three template DNA molecules. (Twelve wells are shown for the purpose of illustration. In practice, greater than 100 wells would be used.) Most wells do not contain a mutant *APC* template. However, wells that do contain a mutant *APC* template will contain relatively few competing wild type *APC* templates. Following amplification of the *APC* genes in each well by PCR, the *APC* open reading frames in the amplified products are first transcribed and then translated in vitro. Synthetic proteins derived from wild type *APC* alleles are full-length, whereas the majority of mutant APC alleles encode truncated proteins. The detection of a truncated *APC* protein in multiple wells by polyacrylamide gel electrophoresis indicates the presence of mutant *APC* alleles – and therefore the presence of tumor cells – in the fecal sample

relatively rare alleles. Instead of numerous DNA sequencing reactions that would be required to test for diverse *APC* mutations, this test employs sequential *in vitro* transcription and translation to produce template-encoded APC proteins.

A similar approach has been employed to detect mismatch instability in fecal DNA. While neither of these tests has yet entered clinical practice, they do illustrate noninvasive approaches to detect tumors by the cancer genes that triggered them.

Bladder cancer. The detection of cancer cells in urine, a technique known as urine cytology, is a common noninvasive procedure for the diagnosis of bladder cancer, but it can miss up to 50% of tumors. The direct visualization and biopsy of suspicious bladder lesions by a technique known as cystoscopy is highly sensitive and specific, but is also invasive, expensive and uncomfortable for the patient. For these reasons, the detection of bladder cancers would be greatly facilitated by a genetic test. Molecular markers of cancer cells that have successfully been detected in the urine of bladder cancer patients include *P53* mutations.

The Majority of Current Anticancer Therapies Inhibit Cell Growth

Most of the anticancer therapies currently in use predate the development of the cancer gene theory. Ionizing radiation and chemotherapeutic drugs that are widely used as both primary and adjuvant forms of therapy were adopted in the clinic not because they necessarily discriminate between normal cells and cancer cells, but because they are potent inhibitors of cell growth.

Many anticancer agents that inhibit cell growth work by one of two general mechanisms and can be thus catagorized:

DNA damaging agents. Double- and single-strand DNA breaks are sensed by the DNA damage signaling network (see Chapter 5). Via multiple downstream signaling pathways, DNA damage triggers growth inhibitory affects such as cell cycle arrest and apoptosis. DNA damaging agents include ionizing radiation and drugs known as *radiomimetics*.

DNA synthesis inhibitors. Because proliferating cell populations replicate their genomic DNA once per cell cycle, inhibition of DNA replication effectively halts cell growth. There are two ways in which DNA synthesis can be inhibited: (1) *nucleotide analogs* of different kinds can either terminate nascent DNA strands or competitively inhibit DNA polymerases, and (2) *antimetabolites* function to inhibit the enzymes that catalyze the synthesis of nucleotides. Effective targets of antimetabolite inhibition include *ribonucleotide reductase* and *thymidylate synthase*. Importantly, inhibition of DNA synthesis can eventually lead to the accumulation of DNA strand breaks. Thus, DNA synthesis inhibitors can indirectly trigger the DNA damage signaling network.

Anticancer therapy based solely on growth inhibition is often highly successful. The reason behind this success is not obvious. The cells that compose most tumors do not necessarily proliferate at a higher rate than those in normal regenerative

tissue compartments, and in fact tumor cells may proliferate at a lower rate (see Chapter 1). Furthermore, the effects of DNA damaging agents and DNA synthesis inhibitors on DNA are not fundamentally different in normal and tumor cells, nor do these agents interact with cancer genes or the proteins they encode. Yet, despite their non-specificity, these widely used drugs can be highly effective in killing tumor cells and reducing the burden of cancer.

DNA damaging agents and DNA synthesis inhibitors cause chromosome breaks and DNA replication intermediates, respectively, in cancer cells and normal cells alike. The difference lies in the cellular responses to these insults. The genetically programmed responses of cancer cells to aberrant chromosome structures are often defective (see Chapter 5). *p53* is a common node in signaling pathways that monitor chromosome integrity. Loss of p53 function, acquired during tumorigenesis, can decrease a cell's capacity to undergo growth arrest in response to DNA damage and DNA replication intermediates. Analysis of cultured p53-deficient cancer cells exposed to common therapeutic agents has revealed that failure to normally arrest cell cycle progression can cause aberrant cell division, leading to cell death. Analysis of p53-dependent phenotypes has revealed that the genetic alterations that liberate cancer cells from the normal restraints on growth can also leave them uniquely vulnerable to therapeutic agents.

Molecularly Targeted Therapy: *BCR-ABL* and Imatinib

While some types of cancer are exquisitely sensitive to commonly employed forms of anticancer therapy – and are therefore curable or treatable – many cancers remain highly refractive to DNA damaging agents and DNA synthesis inhibitors. New therapeutic strategies are desperately needed. Among the many applications of cancer genetics, perhaps none is more exciting than the use of recurrent genetic alterations to guide the development of new drugs.

The cancer that serves as the best paradigm for gene-based, rational design of anticancer therapy is chronic myelogenous leukemia (CML). CML is a cancer that has, until recently, been difficult to treat. Like many cancers, CML evolves through a series of discrete stages, during which cancer clones progressively accumulate genetic alterations. A stable, or chronic, phase of the disease is characterized by excess numbers of myeloid cells that differentiate normally. Within 4–6 years, the disease passes through an accelerated stage and then enters a terminal stage known as *blast crisis*. Blast crisis is an acute leukemia that is refractory to treatment and invariably fatal.

More than 95% of CML cases exhibit the reciprocal translocation between chromosomes 9 and 22 that creates the *BCR-ABL* oncogene (see Chapter 2). The *BCR-ABL* fusion protein is constitutively expressed and as a result, the tyrosine kinase encoded by *ABL* is highly active in CML cells. Dysregulated ABL activity causes the cancer phenotype of CML. Therefore, inhibition of ABL catalytic activity would be predicted to be an effective strategy for CML therapy.

Protein kinases are common components of key signaling pathways that involve cancer gene-encoded proteins (see Chapter 5). Because protein kinases play central roles in cancer, pharmaceutical companies developed numerous specific inhibitors of these diverse enzymes and tested them as potential anticancer agents. One compound isolated and tested was an inhibitor of the platelet-derived growth factor receptor (PDGF-R). This compound, designated *imatinib mesylate* (often referred to simply as imatinib, alternatively known as STI571 and Gleevec) was subsequently found to also inhibit the ABL tyrosine kinase (see Fig. 7.3). It was demonstrated that imatinib could specifically block the proliferation of cells expressing the *BCR-ABL* oncogene. Preclinical results such as these suggested that imatinib might show efficacy in the treatment of patients with CML.

The clinical trials of imatinib, reported in 2001, were a striking success. Nearly all of the *BCR-ABL*-positive CML patients that were in the chronic phase of the disease achieved long-term remission after imatinib therapy. The patients selected for these trials had previously failed other therapeutic regimens, making the rate of response all the more impressive. Even patients in the midst of blast crisis were found to benefit from imatinib therapy, although the majority of these patients experienced eventual recurrence of disease. Unlike other forms of cancer therapy, imatinib use was associated with only minimal toxicity; only a small percentage of the patients in the trial reported adverse effects and these were generally mild in nature. The rate of remission and the low toxicity of imatinib were unprecedented for an experimental cancer drug.

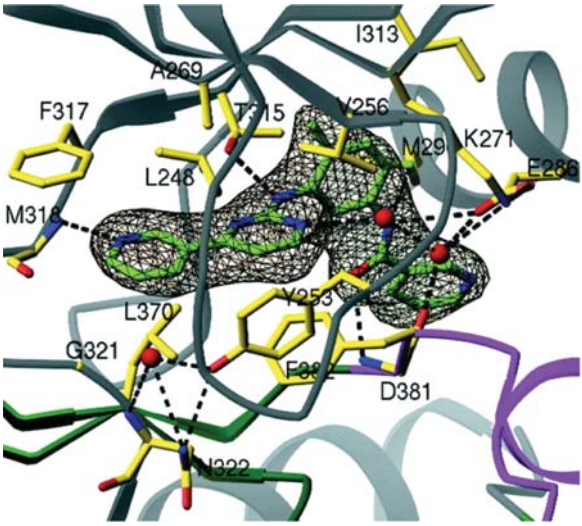

Fig. 7.3 Inhibition of the ABL tyrosine kinase by imatinib. This structural representation demonstrates how the imatinib molecule fits into binds to the nucleotide binding pocket of ABL. Shown are the carbon atoms of the protein (yellow) that interact with the carbon atoms of the imatinib molecule via hydrogen bonds (dashed lines). (Reprinted with permission from Schindler et al. *Science* **289**, 1938–1942. Copyright 2000 AAAS.)

In several respects, CML presents an ideal challenge for molecularly targeted therapy. CML was among the first cancers to be associated with a defined genetic alteration that is nearly universal. The *BCR-ABL* gene is present in the vast majority of CML patients and is the most prominent cause of the cancer phenotype. CML cells require constitutive ABL activity to maintain their highly proliferative state. Not only is BCR-ABL a thoroughly validated target, it is also an enzyme that is inherently 'druggable', that is, a small, diffusible molecular can block ATP-binding and thus inhibit the catalytic moiety. As the clinical trials of imatinib demonstrate, systemic inhibition of ABL kinase activity has little effect on normal proliferating cell populations. One potential explanation for the lack of toxicity is that ABL might function primarily during development and may not be required in adult tissues.

Imatinib will have applications beyond the treatment of CML. In addition to PDGF-R and ABL, imatinib also inhibits the protein tyrosine kinase encoded by the *C-KIT* oncogene. Oncogenic mutations in *C-KIT* drive a relatively rare type of cancer known as the *gastrointestinal stroma tumor* (GIST), a cancer that arises from the mesenchymal tissues of the gut wall. Imatinib treatment of patients with metastatic GISTs has resulted in dramatic regression of disease. C-KIT protein is overexpressed in a fraction of several other tumors, including acute myeloid leukemia, small-cell lung cancer, and melanoma. However, it remains to be established whether *C-KIT* expression is related to tumor cell survival in these cancers.

The successful therapy of CML by imatinib was a watershed event in experimental cancer therapeutics. Most importantly, imatinib provides a powerful treatment for a cancer that recently had been considered incurable. From a research standpoint, imatinib provides a paradigm for the design of specific forms of therapy based on the genetics of a cancer.

The success in treating CML with imatinib will not be easy to replicate in other types of cancer. BCR-ABL is arguably the most well-validated molecular target in cancer. Other cancers have molecular origins that are substantially more diverse than those of CML. Many genetic alterations that give rise to cancer, such as those that cause the loss of function of tumor suppressor genes, are not obviously druggable. Clearly, there are theoretical and practical obstacles to the design of specific therapeutic strategies for some of the most common cancers. Nonetheless, the success of imatinib demonstrates that cancer genes can inform the development of specific approaches to treatment.

Clonal Evolution of Therapeutic Resistance

New therapies such as imatinib are directed against the proteins encoded by targeted cancer genes. The interaction between this drug and its biological target is notable for its specificity. As might have been expected, the clinical responses to imatinib have been found to be closely linked to the original mutation in the target cancer gene but also to secondary mutations that might arise after the initiation of therapy.

The primary mutation in an activating oncogene can largely determine the response to a targeted therapeutic. In patients with gastrointestinal stromal tumors (GISTs), several different somatic *C-KIT* mutations underlie distinct responses to imatinib. Tumors harboring mutations in exon 11 of *C-KIT* are more sensitive to imatinib than are tumors that harbor mutations in *C-KIT* exon 9, for example. As a result, patients with tumors that contain exon 11 *C-KIT* mutations remain disease-free for a longer period and have a greater survival after therapy than those with tumors that express the exon 9 *C-KIT* mutant. Thus, the *C-KIT* alleleotype can be used to predict the initial clinical response of GIST patients to imatinib.

Despite the striking success of imatinib as a therapeutic agent against CML and GIST, many patients eventually become resistant to the effects of the drug and suffer relapse. In such cases, analysis of the target gene, (*BCR-ABL* in CML and *C-KIT* in GIST) often reveals secondary mutations that preserve oncogenic activity but disrupt the inhibitory binding of imatinib. In leukemias, the cells from newly arising clones are mixed with cells from precursor clones and normal cells. For this reason, the process by which secondary mutations develop into drug-resistant cancer is best studied in a solid tumor, such as a GIST, wherein cancer cells grow in clonally derived metastatic foci that can be monitored and sampled.

GISTs that harbor somatically acquired *C-KIT* mutations tend to respond dramatically to the effects of imatinib. However, many patients suffer relapse and develop new metastatic foci with 3 years of treatment. Analysis of these metastatic tumors has revealed a recurrent mutation within the region of *C-KIT* that encodes the first portion of the tyrosine kinase domain. A T→C transition at position 1982 results in an amino acid substitution, V654A. The V654A missense mutation is detected at the time of relapse, on the same allele that harbors the original, primary mutation (see Fig. 7.4). Dual mutations in a single *C-KIT* allele are never found in

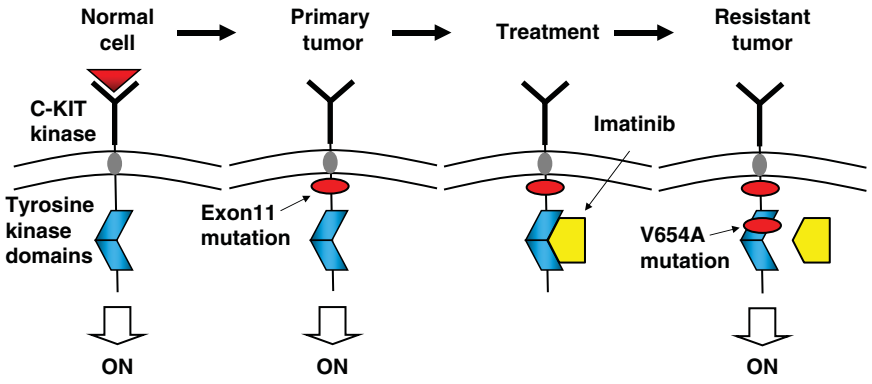

Fig. 7.4 A secondary *C-KIT* mutation causes imatinib resistance. In normal cells, the transmembrane *C-KIT* receptor responds normally to ligand (red triangle). A frequently observed mutation of *C-KIT* within exon 11 causes ligand-independent activation of the *C-KIT* receptor and drives the growth of GISTs. Imatinib (yellow) binds within two tyrosine kinase domains (blue) that span amino acids 598–694 and 771–924, causing a therapeutic response. A secondary mutation affecting codon 654 disrupts the binding of the imatinib molecule, and causes drug-resistant tumor growth

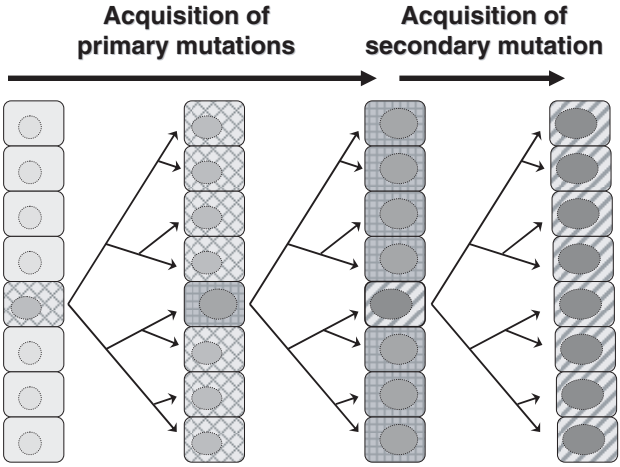

Fig. 7.5 Clonal evolution of drug resistance. Expanding tumors acquire primary mutations by a process consisting of iterative waves of mutation and clonal expansion (see Chapter 1). Drug treatment introduces a new type of selective pressure and thereby drives further clonal evolution. The result is a clone that harbors a secondary mutation which causes drug resistance

patients prior to imatinib therapy. Furthermore, the V654A mutant is only found as a secondary alteration after imatinib therapy, and has never been detected as a primary alteration. Secondary mutations in *C-KIT* within exons 13, 14 and 17 that similarly block the interaction of *C-KIT* with imatinib have also been described.

The process by which neoplastic clones sequentially acquire mutations during tumorigenesis also provides a model for understanding the evolution of drug resistance. Treatment of a cancer with a therapeutic agent causes a new form of selective pressure. As demonstrated by the emergence of imatinib-resistant metastases in GIST, the selective pressure provided by a specific drug can result in the expansion of clones that harbor new mutations in the target gene. Primary mutations arise and are propagated as a result of several rounds of clonal evolution during tumorigenesis. Secondary mutations arise via an additional wave of mutation followed by clonal expansion (see Fig. 7.5).

Allele-specific Cancer Therapy: Gefitinib

Another valid and compelling target for cancer gene-specific therapy is the protein tyrosine kinase encoded by *EGFR*. EGFR activity is dysregulated in several of the most prevalent types of cancer (see Chapter 5). Two highly selective small-molecule inhibitors of the EGFR kinase, named *gefitinib* and *erlotinib*, have been developed by the pharmaceutical industry and tested as anticancer therapeutics.

Non-small-cell lung cancer (NSCLC) is common and often refractory to therapy. These characteristics combine to make it the leading cause of cancer death in the US (see Chapter 6). The majority of NSCLC overexpress EGFR, either as a result of *EGFR* amplification or gain-of-function mutations (see Chapter 5). Clinical trials have tested whether gefitinib, the prototype EGFR inhibitor, might have the

potential to effectively treat metastatic NSCLC, particularly in cases where other forms of therapy have failed.

The initial trial results were mixed. The majority of NSCLC patients did not respond to gefitinib treatment. However, about 10% of patients had responses to gefitinib that were rapid and in many cases dramatic. Interestingly, the patients in the responder group had several identifiable characteristics. Gefitinib responders were disproportionately women, patients who had never smoked, patients with the adenocarcinoma type of NSCLC, and Asians. Genetic analysis revealed that this subgroup had frequent somatic mutations in the *EGFR* gene. The EGFR mutations detected in gefitinib responders included small, in-frame deletions or missense mutations around the domain that encodes the bilobed ATP-binding pocket of the tyrosine kinase moiety (see Fig. 7.6). These mutations cause the repositioning of critical residues that are involved in ATP-binding, thereby stabilizing both the binding of ATP and the binding of gefitinib. Accordingly, *EGFR* mutations that increase EGFR catalytic activity and autophosphorylation simultaneously increase the affinity of EGFR for gefitinib. The affinity of mutant EGFR for gefitinib was unexpected, as this small molecule had originally been designed to inhibit overexpressed, wild type EGFR. Thus, structural studies were able to explain why patients with tumors that harbor *EGFR* mutations responded to gefitinib and those with wild type alleles were less likely to respond.

With respect to gefitinib sensitivity, it appears that not all *EGFR* mutations are equivalent. Cell-based studies have revealed that specific *EGFR* mutations can further predict the sensitivity of cancer cells to gefitinib. The introduction of an exon 20 insertion mutant creates a cancer cell that is 100-fold more resistant to the effects of gefitinib than are the cells expressing the more common deletions in exon 19 and point mutations in exon 21.

Many of the NSCLC patients that initially respond to gefinitib therapy unfortunately go on to develop resistant disease. As was found to be the case in imatinib-treated CML and GIST, a secondary mutation in the target gene (in this case the T790M

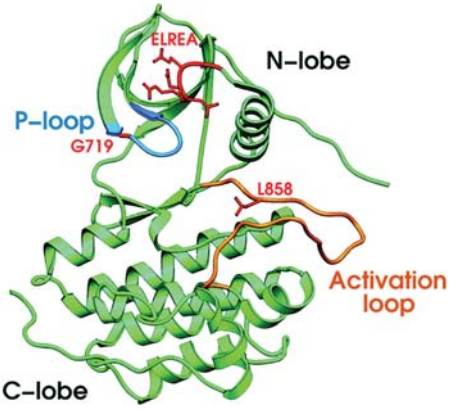

Fig. 7.6 Mutations in *EGFR* sensitize lung cancers to gefitinib. The effects of gefitinib-sensitizing mutations are revealed by the three dimensional structure of the EGFR kinase domain. The two lobes of the kinase domain are as shown. Point mutations affect G719 and alter the P-loop (blue), or L858 in the activation loop (orange). A recurrent in-frame deletion affects the amino acids residues ELREA within the N-lobe. These alterations increase the catalytic activity of EGFR and also increase the affinity of EGFR for gefitinib. (Reprinted with permission from Paez et al. Science **304**, 1497–1500. Copyright 2004 AAAS.)

mutation in *EGFR*) renders new cancer clones resistant to the therapeutic effect of gefinitib. Unlike the V654A mutation in *C-KIT* that causes imatinib resistance, the T790M mutation in *EGFR* can be found as a primary mutation in some patients not treated with gefitinib.

Whereas the secondary mutations that lead to drug resistance present a major problem with molecularly targeted therapy, it also appears that additional allele-specific agents may present a solution. Two different inhibitors of the ABL tyrosine kinase, named *dasatanib* and *nilotinib*, have shown promise in treating CML patients who initially responded to imatinib and subsequently relapsed. These drugs appear to interact with ABL in a slightly different way than does imatinib, and therefore can block the growth of cells harboring *BCR-ABL* alleles that contain a secondary mutation. In a similar manner, inhibitors directed against different structural aspects of the EGFR tyrosine kinase, including molecules designated HKI-272 and EKB-569, are able to inhibit the protein encoded by the T790M *EFGR* mutant. The use of multiple drugs to treat a single neoplasm is called *combination therapy*.

Antibody-Mediated Inhibition of Receptor Tyrosine Kinases

Small-molecule inhibitors have proven effective at blocking the catalytic activity of mutant receptor tyrosine kinases, as illustrated by imatinib and gefitinib. When a receptor tyrosine kinase (RTK) is oncogenically activated by the mechanism of gene amplification, reduction of downstream pathway activation can alternatively be achieved by targeting the extracellular RTK domains, known as *ectodomains*. Therapeutic antibodies directed against ectodomains can either interfere with ligand binding or inhibit receptor dimerization. Both of these strategies result in reduced receptor tyrosine kinase activity, and therefore reduced cell proliferation and survival. Specific monoclonal antibodies have been developed against several RTKs, including the frequently amplified EGFR and ERBB2.

Several forms of antibody therapy have recently been approved for clinical use. *Cetuximab* (also known as C225 or Erbitux) is a therapeutic monoclonal antibody that binds to the ectodomain of EGFR with high affinity. The association of cetuximab with EGFR blocks ligand binding and thus prevents activation of EGFR tyrosine kinase activity. In contrast to EGFR, ERBB2 is a RTK that functions without a ligand, but rather is activated by association with other members of the ERBB family of receptors (see Chapter 5). *Trastuzumab* (also known as Herceptin) is a monoclonal antibody that binds to the extracellular segment of ERBB2 and appears to inhibit the protein–protein interactions that result in ERBB2 activation.

Cetuximab has has shown efficacy in the treatment of some patients with colorectal cancer, head and neck cancer and several other types of solid tumors, while trastuzumab has proven to be useful for the treatment of breast cancers that overexpress ERBB2. Overall, monoclonal antibodies have been found to induce growth arrest and cell death in tumor cells. The efficiency with which antibodies

and small-molecule kinase inhibitors can achieve these effects is roughly similar. For some types of cancers, the combination of monoclonal antibody therapy with small-molecule kinase inhibitors, and also with traditional forms of growth inhibitory therapy, have proven to be synergistic.

One problem that arises with the use of antibodies as drugs are the immune responses triggered by foreign proteins. Monoclonal antibodies typically used for research purposes are most commonly raised in mice. To circumvent problems of cross-species immunogenicity, antibodies used for therapy are engineered to contain protein regions encoded by human genes. Such antibodies are said to be *humanized*. Trastuzumab is an example of a humanized antibody. Cetuximab is a *chimeric* monoclonal antibody, in which the variable regions are derived from mouse genes, while the constant regions of the antibody molecule are derived from human genes.

Targeting Death Receptors: TRAIL

Cell surface proteins known as death receptors trigger apoptosis via the extrinsic pathway (see Chapter 5). The extrinsic pathway of apoptosis is largely independent of the p53 protein, and is therefore intact in the many tumors that harbor *P53*-inactivating mutations.

The tumor necrosis factor (TNF) superfamily of ligands interacts with a large family of cell surface receptors that can regulate both cell proliferation and cell death. A subset of these ligands and receptors preferentially trigger apoptosis pathways. Efforts to therapeutically activate death receptor-mediated apoptosis in cancers have focused on ligands that specifically interact with death receptors.

Not all death receptors are suitable clinical targets. Several of the most common of these are present in normal tissues. The prototypical death-inducing ligand is a protein known as FasL. Because FasL binds Fas receptors that are concentrated in the liver, exogenous administration of FasL would be expected to cause massive necrosis of the liver and thus be highly toxic. In contrast to FasL, the *TNF-related apoptosis inducing ligand* (TRAIL; also known as Apo2L) specifically interacts with several receptors that are less widely distributed, including transmembrane proteins known as death receptors 4 and 5 (DR4 and DR5). In addition, TRAIL binds with at least 2 non-functional receptors that are unable to trigger cell death and are thought to function as decoys. The overall effect of TRAIL is dependent on the relative presence on the cell surface of death receptors and decoy receptors.

Many types of cancer cells have been shown to express significant amounts of DR4 and/or DR5, although the genetic basis for cancer cell-specific expression remains a topic of investigation. Cancer cells that express high levels of death receptors and low levels of decoy receptors tend to respond to TRAIL administration by triggering apoptosis. Soluble, recombinant TRAIL prepared as a pharmaceutical agent has been shown to target a wide range of tumor cell types and appears to have fewer toxic side effects than other death receptor ligands. Early

efforts to combine death receptor-targeted therapy with conventional therapies have shown promise.

The normal physiological role of endogenous TRAIL remains incompletely understood. It is has been suggested that TRAIL may function as part of an immune surveillance mechanism to detect and eliminate oncogene-transformed and virus-infected cells. If this is in fact the case, the use of TRAIL as an anticancer drug would represent an atttractive means of pharmacologically enhancing normal anticancer defenses.

Customized Cancer Therapy

Cancer genes are the cause of cancer, but they may also be the keys that can unlock the cure. Drugs directed against specific targets have demonstrated effectiveness in treating several common cancers that had responded poorly to older modes of therapy. The analysis of cancer genes has even provided insights and possible solutions to treatment failures. These early experiences have generated a great deal of optimism surrounding the feasibility of rationally designed, targeted therapy. The foundation of this new approach to treating cancer patients is an understanding of cancer genetics.

Several simple principles underlie recent efforts to pharmacologically inhibit activated oncogenes:

Recurrent genetic alterations define molecular targets. The successful therapy of CML with imatinib demonstrates that targeting an oncogene-encoded protein required for cell survival can be a highly effective therapeutic strategy. This approach requires both a valid target and a specific inhibitor of that target.

Mutations in target genes can predict therapeutic responses. Because targeted therapy depends upon the specific molecular interaction between a drug and a protein, distinct mutations within a target gene can affect efficacy. This principle is vividly illustrated by the mutations in *EGFR* that affect responses to gefitinib.

Secondary mutations cause the development of therapeutic resistance. Targeted therapeutics create selective pressure that can drive further clonal evolution. Secondary mutations that prevent inhibitor binding but preserve oncoprotein function provide a significant selective advantage. Acquired resistance to both imatinib and gefitinib have been attributed to secondary mutations.

Combination therapy can overcome resistance. Clonal evolution in essence creates a moving target. Fortunately, the use of multiple agents that interact with a single target in different ways can circumvent this problem. Drugs that interact with distinct target molecules can also be combined to enhance efficacy. As cancers evolve, so do the therapeutic strategies to defeat them.

The use of highly specific tyrosine kinase inhibitors to treat cancers with defined genetic alterations is a significant departure from more general growth inhibitory strategies that predate the cancer gene theory. Older therapies that induce DNA damage and block DNA replication, often combined with surgery, are the current

mainstays of therapy for the majority of cancers and will continue to be important for the foreseeable future. However, continued improvements in cancer survival will likely emerge from the combined use of established therapies and new, targeted drugs.

In the future, genetic information will play a larger role in treatment planning. The ability to accurately predict an individual's response to therapy based germline and somatically acquired cancer genes signature may one day allow the formulation of a customized course of therapy, optimized for each patient.

Further Reading

Azam, M., Latek, R. R. & Daley, G. Q. Mechanisms of autoinhibition and STI-571/imatinib resistance revealed by mutagenesis of BCR-ABL. *Cell* **112**, 831–843 (2003).

Domchek, S. M. & Weber, B. L. Clinical management of BRCA1 and BRCA2 mutation carriers. *Oncogene* **25**, 5825–5831 (2006).

Druker, B. J. Perspectives on the development of a molecularly targeted agent. *Cancer Cell* **1**, 31–36 (2002).

Greulich, H. et al. Oncogenic transformation by inhibitor-sensitive and -resistant EGFR mutants. *PLoS Med*. **2**, e313 (2005).

Guttmacher, A. E. & Collins, F. S. Realizing the promise of genomics in biomedical research. *JAMA* **294**, 1399–1402 (2005).

Herbst, R. S., Fukuoka, M. & Baselga, J. Gefitinib – a novel targeted approach to treating cancer. *Nat. Rev. Cancer* **4**, 956–965 (2004).

Hu, Y. C., Sidransky, D. & Ahrendt, S. A. Molecular detection approaches for smoking associated tumors. *Oncogene* **21**, 7289–7297 (2002).

Hynes, N. E. & Lane, H. A. ERBB receptors and cancer: The complexity of targeted inhibitors. *Nat. Rev. Cancer* **5**, 341–354 (2005).

Kelley, S. K. & Ashkenazi, A. Targeting death receptors in cancer with Apo2L/TRAIL. *Curr. Opin. Pharmacol*. **4**, 333–339 (2004).

Krause, D. S. & Van Etten, R. A. Tyrosine kinases as targets for cancer therapy. *N. Engl. J. Med*. **353**, 172–187 (2005).

Lacroix, M. Significance, detection and markers of disseminated breast cancer cells. *Endocr. Relat. Cancer* **13**, 1033–1067 (2006).

Mao, L. et al. Microsatellite alterations as clonal markers for the detection of human cancer. *Proc. Natl. Acad. Sci. U. S. A*. **91**, 9871–9875 (1994).

Mills, N. E. et al. Detection of K-ras oncogene mutations in bronchoalveolar lavage fluid for lung cancer diagnosis. *J. Natl. Cancer Inst*. **87**, 1056–1060 (1995).

Morgensztern, D. & Govindan, R. Is there a role for cetuximab in non small cell lung cancer? *Clin. Cancer Res*. **13**, 4602s–4605s (2007).

Nahta, R. & Esteva, F. J. Trastuzumab: Triumphs and tribulations. *Oncogene* **26**, 3637–3643 (2007).

Petitjean, A., Achatz, M. I., Borresen-Dale, A. L., Hainaut, P. & Olivier, M. TP53 mutations in human cancers: Functional selection and impact on cancer prognosis and outcomes. *Oncogene* **26**, 2157–2165 (2007).

Schindler, T. et al. Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science* **289**, 1938–1942 (2000).

Schwartz, R. S. A needle in a haystack of genes. *N. Engl. J. Med*. **346**, 302–304 (2002).

Sharma, S. V., Bell, D. W., Settleman, J. & Haber, D. A. Epidermal growth factor receptor mutations in lung cancer. *Nat. Rev. Cancer* **7**, 169–181 (2007).

Trepanier, A. et al. Genetic cancer risk assessment and counseling: Recommendations of the national society of genetic counselors. *J. Genet. Couns*. **13**, 83–114 (2004).

Wang, S. & El-Deiry, W. S. TRAIL and apoptosis induction by TNF-family death receptors. *Oncogene* **22**, 8628–8633 (2003).

Wexler, N. S. The Tiresias complex: Huntington's disease as a paradigm of testing for late-onset disorders. *FASEB J*. **6**, 2820–2825 (1992).

Appendix

A Catalog of Cancer Genes

An extensive compilation of confirmed cancer genes can be found in the Catalog of Somatic Mutations in Cancer (COSMIC), an online database maintained by the Sanger Institute (Forbes et al. 2006). This database provides a sense of the number and diversity of cancer genes and their functional scope. Many additional cancer gene mutations have been discovered via high throughput sequencing of cancer genomes.

Further Reading

Forbes, S., Clements, J., Dawson, E., Bamford, S., Webb, T., Dogan, A., Flanagan, A., Teague, J., Wooster, R., Futreal, P. A. & Stratton, M. R. COSMIC 2005. *Br J Cancer* **94**, 318–322 (2006). http://www.sanger.ac.uk/genetics/CGP/cosmic/

Futreal, P. A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., & Stratton, M. R. A census of human cancer genes. *Nat Rev Cancer* **4**, 177–183 (2004).

Symbol	Name	Location	Mutations		Tumor types (somatic mutations)*	Tumor types (germline mutations)*	Cancer syndrome	Mode	Mutation type(s)*
			Somatic	Germline					
ABL1	v-abl Abelson murine leukemia viral oncogene homolog 1	9q34.1	Yes		CML, ALL, T-ALL			Dom	T, Mis
ABL2	v-abl Abelson murine leukemia viral oncogene homolog 2	1q24-q25	Yes		AML			Dom	T
AF15Q14	AF15q14 protein	15q14	Yes		AML			Dom	T
AF1Q	ALL1-fused gene from chromosome 1q	1q21	Yes		ALL			Dom	T
AF3p21	SH3 protein interacting with Nck, 90 kDa (ALL1-fused gene from 3p21)	3p21	Yes		ALL			Dom	T
AF5q31	ALL1-fused gene from 5q31	5q31	Yes		ALL			Dom	T
AKAP9	A kinase (PRKA) anchor protein (yotiao) 9	7q21-q22	Yes		Papillary thyroid			Dom	T
AKT2	v-akt murine thymoma viral oncogene homolog 2	19q13.1-q13.2	Yes		Ovarian, pancreatic			Dom	A
ALK	Anaplastic lymphoma kinase (Ki-1)	2p23	Yes		ALCL			Dom	T
ALO17	KIAA1618 protein	17q25.3	Yes		ALCL			Dom	T

APC	Adenomatous polyposis of the colon gene	5q21	Yes	Yes	Colorectal, pancreatic, desmoid, hepatoblastoma, glioma, other CNS	Colorectal, pancreatic, desmoid, hepatoblastoma, glioma, other CNS	Adenomatous polyposis coli; Turcot syndrome	Rec	D, Mis, N, F, S
ARHGEF12	RHO guanine nucleotide exchange factor (GEF) 12 (LARG)	11q23.3	Yes		AML			Dom	T
ARHH	RAS homolog gene family, member H (TTF)	4p13	Yes		NHL			Dom	T
ARNT	Aryl hydrocarbon receptor nuclear translocator	1q21	Yes		AML			Dom	T
ASPSCR1	Alveolar soft part sarcoma chromosome region, candidate 1	17q25	Yes		Alveolar soft part sarcoma			Dom	T
ATF1	Activating transcription factor 1	12q13	Yes		Malignant melanoma of soft parts, angiomatoid fibrous histiocytoma			Dom	T
ATIC	5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase	2q35	Yes		ALCL			Dom	T
ATM	Ataxia telangiectasia mutated	11q22.3	Yes	Yes	T-PLL	Leukemia, lymphoma, medulloblastoma glioma	Ataxia-telangiectasia	Rec	D, Mis, N, F, S

(continued)

Symbol	Name	Location	Mutations		Tumor types (somatic mutations)*	Tumor types (germline mutations)*	Cancer syndrome	Mode	Mutation type(s)*
			Somatic	Germline					
BCL10	B-cell CLL/lymphoma 10	1p22	Yes		MALT			Dom	T
BCL11A	B-cell CLL/lymphoma 11A	2p13	Yes		B-CLL			Dom	T
BCL11B	B-cell CLL/lymphoma 11B (CTIP2)	14q32.1	Yes		T-ALL			Dom	T
BCL2	B-cell CLL/lymphoma 2	18q21.3	Yes		NHL, CLL			Dom	T
BCL3	B-cell CLL/lymphoma 3	19q13	Yes		CLL			Dom	T
BCL5	B-cell CLL/lymphoma 5	17q22	Yes		CLL			Dom	T
BCL6	B-cell CLL/lymphoma 6	3q27	Yes		NHL, CLL			Dom	T, Mis
BCL7A	B-cell CLL/lymphoma 7A	12q24.1	Yes		BNHL			Dom	T
BCL9	B-cell CLL/lymphoma 9	1q21	Yes		B-ALL			Dom	T
BCR	Break-point cluster region	22q11.21	Yes		CML, ALL			Dom	T
BHD	Folliculin, Birt-Hogg–Dube syndrome	17p11.2		Yes		Renal, fibrofol-liculomas trichodis-comas	Birt-Hogg-Dube syndrome	Rec?	Mis, N, F
BIRC3	Baculoviral IAP repeat-containing 3	11q22-q23	Yes		MALT			Dom	T
BLM	Bloom syndrome	15q26.1		Yes		Leukemia, lym-phoma, skin squamous cell, other cancers	Bloom s yndrome	Rec	Mis, N, F

BMPR1A	Bone morphogenetic protein receptor, type IA	10q22.3		Yes		Gastrointestinal polyps	Juvenile polyposis	Rec	Mis, N, F
BRAF	v-raf murine sarcoma viral oncogene homolog B1	7q34	Yes		Melanoma, colorectal, papillary thyroid, borderline ov, non small-cell lung cancer (NSCLC), cholangiocarcinoma			Dom	Mis, T
BRCA1	Familial breast/ovarian cancer gene 1	17q21	Yes	Yes	Ovarian	Breast, ovarian	Hereditary breast/ovarian cancer	Rec	D, Mis, N, F, S
BRCA2	Familial breast/ovarian cancer gene 2	13q12	Yes	Yes	Breast, ovarian, pancreatic	Breast, ovarian, pancreatic, leukemia (FANCB, FANCD1)	Hereditary breast/ovarian cancer	Rec	D, Mis, N, F, S
BRD4	Bromodomain containing 4	19p13.1	Yes		Lethal midline carcinoma of young people			Dom	T
BRIP1	BRCA1-interacting protein C-terminal helicase 1	17q22		Yes		AML, leukemia, breast	Fanconi anaemia J, breast cancer susceptiblity	Rec	F, N, Mis
BTG1	B-cell translocation gene 1, anti-proliferative	12q22	Yes		BCLL			Dom	T
BUB1B	BUB1 budding uninhibited by benzimidazoles 1 homolog beta	15q15		Yes		Rhabdomyosarcoma	Mosaic variegated aneuploidy	Rec	Mis, N, F, S

(continued)

Symbol	Name	Location	Mutations		Tumor types (somatic mutations)*	Tumor types (germline mutations)*	Cancer syndrome	Mode	Mutation type(s)*
			Somatic	Germline					
C12orf9	Chromosome 12 open reading frame 9	12q14.3	Yes		Lipoma			Dom	T
CARS	Cysteinyl-tRNA synthetase	11p15.5	Yes		ALCL			Dom	T
CBFA2T1	Core-binding factor, runt domain, alpha subunit 2;translocated to, 1	8q22	Yes		AML			Dom	T
CBFA2T3	Core-binding factor, runt domain, alpha subunit 2; translocated to, 3	16q24	Yes		AML			Dom	T
CBFB	Core-binding factor, beta subunit	16q22	Yes		AML			Dom	T
CBL	Cas-Br-M ecotropic retroviral transforming	11q23.3	Yes		AML			Dom	T
CCND1	Cyclin D1	11q13	Yes		CLL, B-ALL, breast			Dom	T
CCND2	Cyclin D2	12p13	Yes		NHL,CLL			Dom	T
CCND3	Cyclin D3	6p21	Yes		MM			Dom	T
CDH1	Cadherin 1, type 1, E-cadherin (ECAD)	16q22.1	Yes	Yes	Lobular breast, gastric	Gastric	Familial gastric carcinoma	Rec	Mis, N, F, S
CDH11	Cadherin 11, type 2, OB-cadherin	16q22.1	Yes		Aneurysmal bone cysts			Dom	T

CDK4	Cyclin-dependent kinase 4	12q14		Yes		Melanoma	Familial malignant melanoma	Dom	Mis
CDK6	Cyclin-dependent kinase 6	7q21-q22	Yes		ALL			Dom	T
CDKN2A-p14ARF	Cyclin-dependent kinase inhibitor 2A- p14ARF protein	9p21	Yes	Yes	Melanoma, multiple other tumour types	Melanoma, pancreatic	Familial malignant melanoma	Rec	D, S
CDKN2A-p16(INK4a)	Cyclin-dependent kinase inhibitor 2A-(p16(INK4a)) gene	9p21	Yes	Yes	Melanoma, multiple other tumour types	Melanoma, pancreatic	Familial malignant melanoma	Rec	D, Mis, N, F, S
CDX2	Caudal type-homeo box transcription factor 2	13q12.3	Yes		AML			Dom	T
CEBPA	CCAAT-/enhancer-binding protein (C/EBP), alpha	11p15.5	Yes		AML, MDS			Dom	Mis, N, F
CEP1	Centrosomal protein 1	9q33	Yes		MPD, NHL			Dom	T
CHK2	CHK2 checkpoint homolog	22q12.1		Yes		Breast	Familial breast cancer	Rec	F
CHIC2	Cysteine-rich hydrophobic domain 2	4q11-q12	Yes		AML			Dom	T
CHN1	Chimerin (chimaerin) 1	2q31-q32.1	Yes		Extraskeletal myxoid chond-rosarcoma			Dom	T
CIC	Capicua homolog	19q13.2	Yes		Soft tissue sarcoma			Dom	T
CLTC	Clathrin, heavy polypeptide (Hc)	17q11-qter	Yes		ALCL			Dom	T
CLTCL1	Clathrin, heavy polypeptide-like 1	22q11.21	Yes		ALCL			Dom	T

(continued)

Symbol	Name	Location	Mutations		Tumor types (somatic mutations)*	Tumor types (germline mutations)*	Cancer syndrome	Mode	Mutation type(s)*
			Somatic	Germline					
CMKOR1	Chemokine orphan receptor 1	2q37.3	Yes		Lipoma			Dom	T
COL1A1	Collagen, type I, alpha 1	17q21.31-q22	Yes		Dermatofibrosarcoma protuberans, aneurysmal bone cyst			Dom	T
COPEB	Core promoter element-binding protein (KLF6)	10p15	Yes		Prostate, glioma			Rec	Mis, N
COX6C	Cytochrome c oxidase subunit VIc	8q22-q23	Yes		Uterine leiomyoma			Dom	T
CREB1	cAMP-responsive element-binding protein 1	2q34	Yes		Clear cell sarcoma			Dom	T
CREBBP	CREB-binding protein (CBP)	16p13.3	Yes		AL, AML			Dom	T
CTNNB1	Catenin (cadherin-associated protein), beta 1	3p22-p21.3	Yes		Colorectal, cvarian, hepatoblastoma, others			Dom	H, Mis
CYLD	Familial cylindromatosis gene	16q12-q13	Yes	Yes	Cylindroma	Cylindroma	Familial cylindromatosis	Rec	Mis, N, F, S
D10S170	DNA segment on chromosome 10 (unique) 170, H4 gene (PTC1)	10q21	Yes		Papillary thyroid, CML			Dom	T

DDB2	Damage-specific DNA-binding protein 2	11p12	Yes		Skin basal cell, skin squamous cell, melanoma	Xeroderma pigmentosum (E)	Rec	Mis, N
DDIT3	DNA-damage-inducible transcript 3	12q13.1-q13.2	Yes	Liposarcoma			Dom	T
DDX10	DEAD (Asp-Glu-Ala-Asp) box polypeptide 10	11q22-q23	Yes	AML*			Dom	T
DDX6	DEAD (Asp-Glu-Ala-Asp) box polypeptide 6	11q23.3	Yes	B-NHL			Dom	T
DEK	DEK oncogene (DNA binding)	6p23	Yes	AML			Dom	T
DUX4	Double homeobox, 4	4q35	Yes	Soft tissue sarcoma			Dom	T
EGFR	Epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homolog, avian)	7p12.3-p12.1	Yes	Glioma, NSCLC			Dom	A, O, Mis
EIF4A2	Eukaryotic translation initiation factor 4A, isoform 2	3q27.3	Yes	NHL			Dom	T
ELF4	E74-like factor 4 (ets domain transcription factor)	Xq26	Yes	AML			Dom	T
ELKS	ELKS protein	12p13.3	Yes	Papillary thyroid			Dom	T
ELL	ELL gene (11–19 lysine-rich leukemia gene)	19p13.1	Yes	AL			Dom	T
EP300	300 kd E1A-binding protein gene	22q13	Yes	Colorectal, breast, pancreatic, AML			Rec	T

(continued)

Symbol	Name	Location	Mutations		Tumor types (somatic mutations)*	Tumor types (germline mutations)*	Cancer syndrome	Mode	Mutation type(s)*
			Somatic	Germline					
EPS15	Epidermal growth factor receptor pathway substrate 15 (AF1p)	1p32	Yes		ALL			Dom	T
ERBB2	v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian)	17q21.1	Yes		Breast, ovarian, other tumour types, NSCLC, gastric			Dom	A, Mis, O
ERCC4	Excision repair cross-complementing rodent repair deficiency, complementation group 4	16p13.3-p13.13		Yes		Skin basal cell, skin squamous cell, melanoma	Xeroderma pigmentosum (F)	Rec	Mis, N, F
ERG	v-ets erythroblastosis virus E26 oncogene-like (avian)	21q22.3	Yes		Ewings sarcoma, prostate, AML			Dom	T
ETV1	ets variant gene 1	7p22	Yes		Ewings sarcoma, prostate			Dom	T
ETV4	ets variant gene 4 (E1A enhancer-binding protein, E1AF)	17q21	Yes		Ewings sarcoma			Dom	T
ETV6	ets variant gene 6 (TEL oncogene)	12p13	Yes		Congenital fibrosarcoma, multiple leukemia and lymphoma, secretory breast, MDS			Dom	T

EVI1	Ecotropic viral integration site 1	3q26	Yes	AML, CML			Dom	T
EWSR1	Ewing sarcoma break point region 1 (EWS)	22q12	Yes	Ewings sarcoma, desmoplastic small round cell tumor, ALL, clear cell sarcoma, sarcoma	w		Dom	T
EXT1	Multiple exostoses type 1 gene	8q24.11-q24.13	Yes		Exostoses, osteosarcoma	Multiple exostoses type 1	Rec	Mis, N, F, S
EXT2	Multiple exostoses type 2 gene	11p12-p11	Yes		Exostoses, osteosarcoma	Multiple exostoses type 2	Rec	Mis, N, F, S
FACL6	Fatty acid-coenzyme A ligase, long-chain 6	5q31	Yes	AML, AEL			Dom	T
FANCA	Fanconi anemia, complementation group A	16q24.3	Yes		AML, leukemia	Fanconi anaemia A	Rec	D, Mis, N, F, S
FANCC	Fanconi anemia, complementation group C	9q22.3	Yes		AML, leukemia	Fanconi anaemia C	Rec	D, Mis, N, F, S
FANCD2	Fanconi anemia, complementation group D2	3p26	Yes		AML, leukemia	Fanconi anaemia D2	Rec	D, Mis, N, F
FANCE	Fanconi anemia, complementation group E	6p21-p22	Yes		AML, leukemia	Fanconi anaemia E	Rec	N, F, S
FANCF	Fanconi anemia, complementation group F	11p15	Yes		AML, leukemia	Fanconi anaemia F	Rec	N, F
FANCG	Fanconi anemia, complementation group G	9p13	Yes		AML, leukemia	Fanconi anaemia G	Rec	Mis, N, F, S

(continued)

Symbol	Name	Location	Mutations		Tumor types (somatic mutations)*	Tumor types (germline mutations)*	Cancer syndrome	Mode	Mutation type(s)*
			Somatic	Germline					
FBXW7	F-box and WD-40 domain protein 7	4q31.3	Yes		Colorectal, endometrial			Dom	Mis, N
FCGR2B	Fc fragment of IgG, low-affinity IIb, receptor for (CD32)	1q23	Yes		ALL			Dom	T
FEV	FEV protein – (HSRNAFEV)	2q36	Yes		Ewings sarcoma			Dom	T
FGFR1	Fibroblast growth factor receptor 1	8p11.2-p11.1	Yes		MPD, NHL			Dom	T
FGFR1OP	FGFR1 oncogene partner (FOP)	6q27	Yes		MPD, NHL			Dom	T
FGFR2	Fibroblast growth factor receptor 2	10q26	Yes		Gastric			Dom	Mis
FGFR3	Fibroblast growth factor receptor 3	4p16.3	Yes		Bladder, MM, T-cell lymphoma			Dom	Mis, T
FH	Fumarate hydratase	1q42.1		Yes	Lieomyomatosis, renal		Hereditary leiomyomatosis and renal cell cancer	Rec	Mis, N, F
FIP1L1	FIP1 like 1	4q12	Yes		Idiopathic hypereosinophilic syndrome			Dom	T
FLI1	Friend leukemia virus integration 1	11q24	Yes		Ewings sarcoma			Dom	T
FLT3	fms-related tyrosine kinase 3	13q12	Yes		AML, ALL			Dom	Mis, O

FNBP1	Formin-binding protein 1 (FBP17)	9q23	Yes	AML	Dom	T
FOXO1A	Forkhead box O1A (FKHR)	13q14.1	Yes	Alveolar rhabdomyosarcomas	Dom	T
FOXO3A	Forkhead box O3A	6q21	Yes	AL	Dom	T
FSTL3	Follistatin-like 3 (secreted glycoprotein)	19p13	Yes	B-CLL	Dom	T
FUS	Fusion, derived from t(12;16) malignant liposarcoma	16p11.2	Yes	Liposarcoma, AML	Dom	T
FVT1	Follicular lymphoma variant translocation 1	18q21.3	Yes	B-NHL	Dom	T
GAS7	Growth arrest-specific 7	17p	Yes	AML*	Dom	T
GATA1	GATA-binding protein 1	Xp11.23	Yes	Megakaryoblastic leukemia of Downs syndrome	Dom	Mis, F
GMPS	Guanine monphosphate synthetase	3q24	Yes	AML	Dom	T
GNAS	Guanine nucleotide binding protein (G protein), alpha stimulating activity polypeptide 1	20q13.2	Yes	Pituitary adenoma	Dom	Mis
GOLGA5	Golgi autoantigen, golgin subfamily a, 5	14q	Yes	Papillary thyroid	Dom	T
GOPC	Golgi-associated PDZ and coiled-coil motif containing	6q21	Yes	Glioblastoma	Dom	O

(continued)

Symbol	Name	Location	Mutations		Tumor types (somatic mutations)*	Tumor types (germline mutations)*	Cancer syndrome	Mode	Mutation type(s)*
			Somatic	Germline					
GPC3	Glypican 3	Xq26.1		Yes		Wilms tumour	Simpson-Golabi–Behmel syndrome	Rec/X	T, D, Mis, N, F, S
GPHN	Gephyrin (GPH)	14q24	Yes		AL			Dom	T
GRAF	GTPase regulator associated with focal adhesion kinase pp125(FAK)	5q31	Yes		AML, MDS			Dom	T, F, S
HCMOGT-1	Sperm antigen HCMOGT-1	17p11.2	Yes		JMML			Dom	T
HEAB	ATP_GTP-binding protein	11q12	Yes		AML			Dom	T
HEI10	Enhancer of invasion 10 – fused to HMGA2	14q11.1	Yes		Uterine leiomyoma			Dom	T
HIP1	Huntingtin-interacting protein 1	7q11.23	Yes		CMML			Dom	T
HIST1H4I	Histone 1, H4i (H4FM)	6p21.3	Yes		NHL			Dom	T
HLF	Hepatic leukemia factor	17q22	Yes		ALL			Dom	T
HLXB9	Homeo box HB9	7q36	Yes		AML			Dom	T
HMGA1	High-mobility group AT-hook 1	6p21	Yes		Microfollicular thyroid adenoma, various benign mesenchymal tumors			Dom	T
HMGA2	High-mobility group AT-hook 2 (HMGIC)	12q15	Yes		Lipoma			Dom	T

HOXA11	Homeo box A11	7p15-p14.2	Yes		CML			Dom	T
HOXA13	Homeo box A13	7p15-p14.2	Yes		AML			Dom	T
HOXA9	Homeo box A9	7p15-p14.2	Yes		AML*			Dom	T
HOXC11	Homeo box C11	12q13.3	Yes		AML			Dom	T
HOXC13	Homeo box C13	12q13.3	Yes		AML			Dom	T
HOXD11	Homeo box D11	2q31-q32	Yes		AML			Dom	T
HOXD13	Homeo box D13	2q31-q32	Yes		AML*			Dom	T
HRAS	v-Ha-ras Harvey rat sarcoma viral oncogene homolog	11p15.5	Yes	Yes	Infrequent sarcomas, rare other types	Rhadomyo-sarcoma, ganglio-neuro-blastoma, bladder	Costello syndrome	Dom	Mis
HRPT2	Hyperparathyroidism 2	1q21-q31	Yes	Yes	Parathyroid adenoma	Parathyroid adenoma, mulitiple ossifying jaw fibroma	Hyperparathyroidism–jaw tumor syndrome	Rec	Mis, N, F
HSPCA	Heat shock 90kDa protein 1, alpha	1q21.2-q22	Yes		NHL			Dom	T
HSPCB	Heat shock 90kDa protein 1, beta	6p12	Yes		NHL			Dom	T
IGHM	Immunoglobulin heavy locus	14q32.33	Yes		MM, Burkitt lymphoma, NHL, CLL, B-ALL, MALT, MLCLS			Dom	T
IGKC	Immunoglobulin kappa locus	2p12	Yes		Burkitt lymphoma, B-NHL			Dom	T

(continued)

Symbol	Name	Location	Mutations		Tumor types (somatic mutations)*	Tumor types (germline mutations)*	Cancer syndrome	Mode	Mutation type(s)*
			Somatic	Germline					
IGLC1	Immunoglobulin lambda locus	22q11.1-q11.2	Yes		Burkitt lymphoma			Dom	T
IL2	Interleukin 2	4q26-q27	Yes		Intestinal T-cell lymphoma			Dom	T
IL21R	Interleukin 21 receptor	16p11	Yes		NHL			Dom	T
IRF4	Interferon regulatory factor 4	6p25-p23	Yes		MM			Dom	T
IRTA1	Immunoglobulin superfamily receptor translocation associated 1	1q21	Yes		B-NHL			Dom	T
ITK	IL2-inducible T-cell kinase	5q31-q32	Yes		Peripheral T-cell lymphoma			Dom	T
JAK2	Janus kinase 2	9p24	Yes		ALL, AML, MPD			Dom	T, Mis, O
JAZF1	Juxtaposed with another zinc finger gene 1	7p15.2-p15.1	Yes		Endometrial stromal tumours			Dom	T
KIT	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog	4q12	Yes	Yes	GIST, AML, TGCT, mastocytosis	GIST, epithelioma	Familial gastrointestinal stromal tumour	Dom	Mis, O
KRAS	v-Ki-ras2 Kirsten rat sarcoma 2 viral oncogene homolog	12p12.1	Yes		Pancreatic, colorectal, lung, thyroid, AML, others			Dom	Mis
KTN1	Kinectin 1 (kinesin receptor)	14q22.1	Yes		Papillary thryoid			Dom	T

LAF4	Lymphoid nuclear protein related to AF4	2q11.2-q12	Yes	ALL	Dom	T
LASP1	LIM and SH3 protein 1	17q11-q21.3	Yes	AML	Dom	T
LCK	Lymphocyte-specific protein tyrosine kinase	1p35-p34.3	Yes	T-ALL	Dom	T
LCP1	Lymphocyte cytosolic protein 1 (L-plastin)	13q14.1-q14.3	Yes	NHL	Dom	T
LCX	Leukemia-associated protein with a CXXC domain	10q21	Yes	AML	Dom	T
LHFP	Lipoma HMGIC fusion partner	13q12	Yes	Lipoma	Dom	T
LIFR	Leukemia inhibitory factor receptor	5p13-p12	Yes	Salivary adenoma	Dom	T
LMO1	LIM domain only 1 (rhombotin 1) (RBTN1)	11p15	Yes	T-ALL	Dom	T
LMO2	LIM domain only 2	11p13	Yes	T-ALL	Dom	T
LPP	LIM domain containing preferred translocation partner in lipoma	3q28	Yes	Lipoma, leukemia	Dom	T
LYL1	Lymphoblastic leukemia-derived sequence 1	19p13.2-p13.1	Yes	T-ALL	Dom	T
MAF	v-maf musculoaponeurotic fibrosarcoma oncogene homolog	16q22-q23	Yes	MM	Dom	T
MAFB	v-maf musculoaponeurotic fibrosarcoma oncogene homolog B (avian)	20q11.2-q13.1	Yes	MM	Dom	T

(continued)

Symbol	Name	Location	Mutations		Tumor types (somatic mutations)*	Tumor types (germline mutations)*	Cancer syndrome	Mode	Mutation type(s)*
			Somatic	Germline					
MALT1	Mucosa-associated lymphoid tissue lymphoma translocation gene 1	18q21	Yes		MALT			Dom	T
MAML2	Mastermind-like 2 (*Drosophila*)	11q22-q23	Yes		Salivary gland mucoepidermoid			Dom	T
MAP2K4	Mitogen-activated protein kinase kinase 4	17p11.2	Yes		Pancreatic, breast, colorectal			Rec	D, Mis, N
MDS1	Myelodysplasia syndrome 1	3q26	Yes		MDS, AML			Dom	T
MDS2	Myelodysplastic syndrome 2	1p36	Yes		MDS			Dom	T
MECT1	Mucoepidermoid translocated 1	19p13	Yes		Salivary gland mucoepidermoid			Dom	T
MEN1	Multiple endocrine neoplasia type 1 gene	11q13	Yes	Yes	Parathyroid tumors	Parathyroid adenoma, pituitary adenoma, pancreatic islet cell, carcinoid	Multiple endocrine neoplasia Type 1	Rec	D, Mis, N, F, S
MET	met proto-oncogene (hepatocyte growth factor receptor)	7q31	Yes		Papillary renal, head-neck squamous cell	Papillary renal	Familial papillary renal cancer	Dom	Mis

MHC2TA	MHC class II transactivator	16p13	Yes		NHL			Dom	T
MKL1	Megakaryoblastic leukemia (translocation) 1	22q13	Yes		Acute megakaryocytic leukemia			Dom	T
MLF1	Myeloid leukemia factor 1	3q25.1	Yes		AML			Dom	T
MLH1	*Escherichia coli* MutL homolog gene	3p21.3	Yes	Yes	Colorectal, endometrial, ovarian, CNS	Colorectal, endometrial, ovarian, CNS	Hereditary nonpolyposis colorectal cancer, Turcot syndrome	Rec	D, Mis, N, F, S
MLL	Myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, *Drosophila*)	11q23	Yes		AML, ALL			Dom	T, O
MLLT1	Myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, *Drosophila*); translocated to, 1 (ENL)	19p13.3	Yes		AL			Dom	T
MLLT10	Myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, *Drosophila*); translocated to, 10 (AF10)	10p12	Yes		AL			Dom	T

(continued)

Symbol	Name	Location	Mutations		Tumor types (somatic mutations)*	Tumor types (germline mutations)*	Cancer syndrome	Mode	Mutation type(s)*
			Somatic	Germline					
MLLT2	Myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, *Drosophila*); translocated to, 2 (AF4)	4q21	Yes		AL			Dom	T
MLLT3	Myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, *Drosophila*); translocated to, 3 (AF9)	9p22	Yes		ALL			Dom	T
MLLT4	Myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, *Drosophila*); translocated to, 4 (AF6)	6q27	Yes		AL			Dom	T
MLLT6	Myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, *Drosophila*); translocated to, 6 (AF17)	17q21	Yes		AL			Dom	T

MLLT7	Myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, *Drosophila*); translocated to, 7 (AFX1)	Xq13.1	Yes		AL			Dom	T
MN1	Meningioma (disrupted in balanced translocation) 1	22q13	Yes		AML, meningioma			Dom	T
MPL	Myeloproliferative leukemia virus oncogene, thrombopoietin receptor	p34	Yes	Yes	MPD	MPD	Familial essential thrombocythemia	Dom	Mis
MSF	MLL septin-like fusion	17q25	Yes		AML*			Dom	T
MSH2	mutS homolog 2 (*E. coli*)	2p22-p21	Yes	Yes	Colorectal, endometrial, ovarian	Colorectal, endometrial, ovarian	Hereditary non-polyposis colorectal cancer	Rec	D, Mis, N, F, S
MSH6	mutS homolog 6 (*E. coli*)	2p16	Yes	Yes	Colorectal	Colorectal, endometrial, ovarian	Hereditary non-polyposis colorectal cancer	Rec	Mis, N, F, S
MSI2	Musashi homolog 2 (*Drosophila*)	17q23.2	Yes		CML			Dom	T
MSN	Moesin	Xq11.2-q12	Yes		ALCL			Dom	T
MTCP1	Mature T-cell proliferation 1	Xq28	Yes		T-cell prolymphocytic leukemia			Dom	T
MUC1	Mucin 1, transmembrane	1q21	Yes		B-NHL			Dom	T

(continued)

Symbol	Name	Location	Mutations		Tumor types (somatic mutations)*	Tumor types (germline mutations)*	Cancer syndrome	Mode	Mutation type(s)*
			Somatic	Germline					
MUTYH	mutY homolog (*E. coli*)	1p34.3–1p32.1		Yes		Colorectal	Adenomatous polyposis coli	Rec	Mis
MYC	v-myc myelocytomatosis viral oncogene homolog (avian)	8q24.12-q24.13	Yes		Burkitt lymphoma, amplified in other cancers, B-CLL			Dom	A, T
MYCL1	v-myc myelocytomatosis viral oncogene homolog 1, lung carcinoma derived (avian)	1p34.3	Yes		Small cell lung			Dom	A
MYCN	v-myc myelocytomatosis viral-related oncogene, neuroblastoma derived (avian)	2p24.1	Yes		Neuroblastoma			Dom	A
MYH11	Myosin, heavy polypeptide 11, smooth muscle	16p13.13-p13.12	Yes		AML			Dom	T
MYH9	Myosin, heavy polypeptide 9, non-muscle	22q13.1	Yes		ALCL			Dom	T
MYST4	MYST histone acetyltransferase (monocytic leukemia) 4 (MORF)	10q22	Yes		AML			Dom	T
NACA	Nascent-polypeptide-associated complex alpha polypeptide	12q23-q24.1	Yes		NHL			Dom	T

NBS1	Nijmegen breakage syndrome 1 (nibrin)	8q21		Yes		NHL, glioma, medullo-blastoma, rhabdomy-osarcoma	Nijmegen breakage syndrome	Rec	Mis, N, F
NCOA2	Nuclear receptor coacti-vator 2 (TIF2)	8q13.1	Yes		AML			Dom	T
NCOA4	Nuclear receptor coactivator 4 - PTC3 (ELE1)	10q11.2	Yes		Papillary thyroid			Dom	T
NF1	Neurofibromatosis type 1 gene	17q12	Yes	Yes	Neurofibroma, glioma	Neurofibroma, glioma	Neurofibro-matosis type 1	Rec	D, Mis, N, F, S, O
NF2	Neurofibromatosis type 2 gene	22q12.2	Yes	Yes	Meningioma, acoustic neuroma	Meningioma, acoustic neuroma	Neurofibro-matosis type 2	Rec	D, Mis, N, F, S, O
NFKB2	Nuclear factor of kappa light polypeptide gene enhancer in B-cells 2 (p49/p100)	10q24	Yes		B-NHL			Dom	T
NIN	Ninein (GSK3B-interact-ing protein)	14q24	Yes		MPD			Dom	T
NONO	Non-POU domain con-taining, octamer-binding	Xq13.1	Yes		Papillary renal cancer			Dom	T
NOTCH1	Notch homolog 1, trans-location-associated (*Drosophila*) (TAN1)	9q34.3	Yes		T-ALL			Dom	T, Mis, O
NPM1	Nucleophosmin (nucleo-lar phosphoprotein B23, numatrin)	5q35	Yes		NHL, APL, AML			Dom	T, F

(continued)

Symbol	Name	Location	Mutations		Tumor types (somatic mutations)*	Tumor types (germline mutations)*	Cancer syndrome	Mode	Mutation type(s)*
			Somatic	Germline					
NR4A3	Nuclear receptor subfamily 4, group A, member 3 (NOR1)	9q22	Yes		Extraskeletal myxoid chondrosarcoma			Dom	T
NRAS	Neuroblastoma RAS viral (v-ras) oncogene homolog	1p13.2	Yes		Melanoma, MM, AML, thyroid			Dom	Mis
NSD1	Nuclear receptor-binding SET domain protein 1	5q35	Yes		AML			Dom	T
NTRK1	Neurotrophic tyrosine kinase, receptor, type 1	1q21-q22	Yes		Papillary thyroid			Dom	T
NTRK3	Neurotrophic tyrosine kinase, receptor, type 3	15q25	Yes		Congenital fibrosarcoma, Secretory breast			Dom	T
NUMA1	Nuclear mitotic apparatus protein 1	11q13	Yes		APL			Dom	T
NUP214	Nucleoporin 214 kDa (CAN)	9q34.1	Yes		AML, T-ALL			Dom	T
NUP98	Nucleoporin 98 kDa	11p15	Yes		AML			Dom	T
NUT	Nuclear protien in testis	q13	Yes		Lethal midline carcinoma of young people			Dom	T
OLIG2	Oligodendrocyte lineage transcription factor 2 (BHLHB1)	21q22.11	Yes		T-ALL			Dom	T
OMD	Osteomodulin	9q22.31	Yes		Aneurysmal bone cysts			Dom	T

PAFAH1B2	Platelet-activating factor acetylhydrolase, isoform Ib, beta subunit 30 kDa	11q23	Yes		MLCLS			Dom	T
PALB2	Partner and localizer of BRCA2	16p12.1		Yes		Wilms tumor, medulloblastoma, AML, breast	Fanconi anaemia N, breast cancer susceptibility	Rec	F, N, Mis
PAX3	Paired box gene 3	2q35	Yes		Alveolar rhabdomyosarcoma			Dom	T
PAX5	Paired box gene 5 (B-cell lineage-specific activator protein)	9p13	Yes		NHL			Dom	T
PAX7	Paired box gene 7	1p36.2-p36.12	Yes		Alveolar rhabdomyosarcoma			Dom	T
PAX8	Paired box gene 8	2q12-q14	Yes		Follicular thyroid			Dom	T
PBX1	Pre-B-cell leukemia transcription factor 1	1q23	Yes		Pre-B-ALL			Dom	T
PCM1	Pericentriolar material 1 (PTC4)	8p22-p21.3	Yes		Papillary thyroid			Dom	T
PCSK7	Proprotein convertase subtilisin/kexin type 7	11q23.3	Yes		MLCLS			Dom	T
PDE4DIP	Phosphodiesterase 4D-interacting protein (myomegalin)	1q12	Yes		MPD			Dom	T
PDGFB	Platelet-derived growth factor beta polypeptide (simian sarcoma viral (v-sis) oncogene homolog)	22q12.3-q13.1	Yes		DFSP			Dom	T

(continued)

Symbol	Name	Location	Mutations		Tumor types (somatic mutations)*	Tumor types (germline mutations)*	Cancer syndrome	Mode	Mutation type(s)*
			Somatic	Germline					
PDGFRA	Platelet-derived growth factor, alpha receptor	4q11-q13	Yes		GIST, idiopathic hypereosinophilic syndrome			Dom	Mis, O, T
PDGFRB	Platelet-derived growth factor receptor, beta polypeptide	5q31-q32	Yes		MPD, AML, CMML, CML			Dom	T
PER1	Period homolog 1 (*Drosophila*)	17p13.1–17p12	Yes		AML, CMML			Dom	T
PHOX2B	Paired-like homeobox 2b	4p12	Yes	Yes	Neuroblastoma	Neuroblastoma	Familial neuroblastoma	Rec	Mis, F
PICALM	Phosphatidylinositol-binding clathrin assembly protein (CALM)	11q14	Yes		TALL, AML,			Dom	T
PIK3CA	Phosphoinositide-3-kinase, catalytic, alpha polypeptide	3q26.3	Yes		Colorectal, gastric, gliobastoma, breast			Dom	Mis
PIM1	pim-1 oncogene	6p21.2	Yes		NHL			Dom	T
PLAG1	Peiomorphic adenoma gene 1	8q12	Yes		Salivary adenoma			Dom	T
PML	Promyelocytic leukemia	15q22	Yes		APL			Dom	T
PMS1	PMS1 postmeiotic segregation increased 1 (*S. cerevisiae*)	2q31-q33		Yes		Colorectal, endometrial, ovarian	Hereditary nonpolyposis colorectal cancer	Rec	Mis, N

PMS2	PMS2 postmeiotic segregation increased 2 (*S. cerevisiae*)	7p22		Yes		Colorectal, endometrial, ovarian, medulloblastoma, glioma	Hereditary nonpolyposis colorectal cancer, Turcot syndrome	Rec	Mis, N, F
PMX1	Paired mesoderm homeo box 1	1q24	Yes		AML			Dom	T
PNUTL1	Peanut-like 1 (*Drosophila*)	22q11.2	Yes		AML			Dom	T
POU2AF1	POU domain, class 2, associating factor 1 (OBF1)	11q23.1	Yes		NHL			Dom	T
POU5F1	POU domain, class 5, transcription factor 1	6p21.31	Yes		Sarcoma			Dom	T
PPARG	Peroxisome proliferative activated receptor, gamma	3p25	Yes		Follicular thyroid			Dom	T
PRCC	Papillary renal cell carcinoma (translocation-associated)	1q21.1	Yes		Papillary renal			Dom	T
PRDM16	PR domain containing 16	1p36.23-p33	Yes		MDS, AML			Dom	T
PRKAR1A	Protein kinase, cAMP-dependent, regulatory, type I, alpha (tissue-specific extinguisher 1)	17q23-q24	Yes	Yes	Papillary thyroid	Myxoma, endocrine, papillary thyroid	Carney complex	Dom, Rec	T, Mis, N, F, S
PRO1073	PRO1073 protein (ALPHA)	11q31.1	Yes		Renal cell carcinoma (childhood epithelioid)			Dom	T

(continued)

Symbol	Name	Location	Mutations		Tumor types (somatic mutations)*	Tumor types (germline mutations)*	Cancer syndrome	Mode	Mutation type(s)*
			Somatic	Germline					
PSIP2	PC4- and SFRS1-interacting protein 2 (LEDGF)	9p22.2	Yes		AML			Dom	T
PTCH	Homolog of *Drosophila* patched gene	9q22.3	Yes	Yes	Skin basal cell, medulloblastoma	Skin basal cell, medulloblastoma	Nevoid basal cell carcinoma syndrome	Rec	Mis, N, F, S
PTEN	Phosphatase and tensin homolog gene	10q23.3	Yes	Yes	Glioma, prostate, endometrial	Harmartoma, glioma, prostate, endometrial	Cowden syndrome, Bannayan-Riley–Ruvalcaba syndrome	Rec	D, Mis, N, F, S
PTPN11	Protein tyrosine phosphatase, non-receptor type 11	12q24.1	Yes		JMML, AML, MDS			Dom	Mis
RAB5EP	Rabaptin, RAB GTPase-binding effector protein 1 (RABPT5)	17p13	Yes		CMML			Dom	T
RAD51L1	RAD51-like 1 (*S. cerevisiae*) (RAD51B)	14q23-q24.2	Yes		Lipoma, uterine leiomyoma			Dom	T
RANBP17	RAN-binding protein 17	5q34	Yes		ALL			Dom	T
RAP1GDS1	RAP1, GTP-GDP dissociation stimulator 1	4q21-q25	Yes		T-ALL			Dom	T
RARA	Retinoic acid receptor, alpha	17q12	Yes		APL			Dom	T

RB1	Retinoblastoma gene	13q14	Yes	Yes	Retinoblastoma, sarcoma, breast, small cell lung	Retinoblastoma, sarcoma, breast, small cell lung	Familial retinoblastoma	Rec	D, Mis, N, F, S
RBM15	RNA-binding motif protein 15	1p13	Yes		Acute megakaryocytic leukemia			Dom	T
RECQL4	RecQ protein-like 4	8q24.3		Yes		Osteosarcoma, skin basal and sqamous cell	Rothmund-Thompson syndrome	Rec	N, F, S
REL	v-rel reticuloendotheliosis viral oncogene homolog (avian)	2p13-p12	Yes		Hodgkin lymphoma			Dom	A
RET	ret proto-oncogene	10q11.2	Yes	Yes	Medullary thyroid, papillary thyroid, pheochromocytoma	Medullary thyroid, papillary thyroid, pheochromocytoma	Multiple endocrine neoplasia 2A/2B	Dom	T, Mis, N, F
ROS1	v-ros UR2 sarcoma virus oncogene homolog 1 (avian)	6q22	Yes		Glioblastoma			Dom	T
RPL22	Ribosomal protein L22 (EAP)	3q26	Yes		AML, CML			Dom	T
RPN1	Ribophorin I	3q21.3-q25.2	Yes		AML			Dom	T
RUNX1	Runt-related transcription factor 1 (AML1)	21q22.3	Yes		AML, pre-B- ALL			Dom	T
RUNXBP2	Runt-related transcription factor-binding protein 2 (MOZ/ZNF220)	8p11	Yes		AML			Dom	T

(continued)

Symbol	Name	Location	Mutations		Tumor types (somatic mutations)*	Tumor types (germline mutations)*	Cancer syndrome	Mode	Mutation type(s)*
			Somatic	Germline					
SBDS	Shwachman–Bodian–Diamond syndrome protein	7q11		Yes		AML, MDS	Schwachman-Diamond syndrome	Rec	Gene Conversion
SDHB	Succinate dehydrogenase complex, subunit B, iron sulfur (Ip)	1p36.1-p35		Yes		Paraganglioma, pheochromocytoma	Familial paraganglioma	Rec	Mis, N, F
SDHC	Succinate dehydrogenase complex, subunit C, integral membrane protein, 15 kDa	1q21		Yes		Paraganglioma, pheochromocytoma	Familial paraganglioma	Rec	Mis, N, F
SDHD	Succinate dehydrogenase complex, subunit D, integral membrane protein	11q23		Yes		Paraganglioma, pheochromocytoma	Familial paraganglioma	Rec	Mis, N, F, S
SEPT6	Septin 6	Xq24	Yes		AML			Dom	T
SET	SET translocation	9q34	Yes		AML			Dom	T
SFPQ	Splicing factor proline/glutamine rich (polypyrimidine tract-binding protein associated)	1p34.3	Yes		Papillary renal cell			Dom	T
SFRS3	Splicing factor, arginine/serine-rich 3	6p21	Yes		Follicular lymphoma			Dom	T
SH3GL1	SH3-domain GRB2-like 1 (EEN)	19p13.3	Yes		AL			Dom	T
SIL	TAL1 (SCL) interrupting locus	1p32	Yes		T-ALL			Dom	T

SMAD4	Homolog of Drosophila mothers against Decapentaplegic 4 gene	18q21.1	Yes	Yes	Colorectal, pancreatic, small intestine	Gastrointestinal polyp	Juvenile polyposis	Rec	D, Mis, N, F
SMARCB1	SWI-/SNF-related, matrix-associated, actin-dependent regulator of chromatin, subfamily b, member 1	22q11	Yes	Yes	Malignant rhabdoid	Malignant rhabdoid	Rhabdoid predisposition syndrome	Rec	D, N, F, S
SMO	Smoothened homolog (*Drosophila*)	7q31-q32	Yes		Skin basal cell			Dom	Mis
SOCS1	Suppressor of cytokine signaling 1	16p13.13	Yes		Hodgkin lymphoma, PMBL			Rec	F, O
SS18	Synovial sarcoma translocation, chromosome 18	18q11.2	Yes		Synovial sarcoma			Dom	T
SS18L1	Synovial sarcoma translocation gene on chromosome 18-like 1	20q13.3	Yes		Synovial sarcoma			Dom	T
SSH3BP1	Spectrin SH3 domain-binding protein 1	10p11.2	Yes		AML			Dom	T
SSX1	Synovial sarcoma, X break point 1	Xp11.23-p11.22	Yes		Synovial sarcoma			Dom	T
SSX2	Synovial sarcoma, X break point 2	Xp11.23-p11.22	Yes		Synovial sarcoma			Dom	T
SSX4	Synovial sarcoma, X break point 4	Xp11.23	Yes		Synovial sarcoma			Dom	T
STK11	Serine/threonine kinase 11 gene (LKB1)	19p13.3	Yes	Yes	NSCLC, pancreatic	Jejunal harmartoma, ovarian, testicular, pancreatic	Peutz–Jeghers syndrome	Rec	D, Mis, N, F, S

(continued)

Symbol	Name	Location	Mutations		Tumor types (somatic mutations)*	Tumor types (germline mutations)*	Cancer syndrome	Mode	Mutation type(s)*
			Somatic	Germline					
STL	Six-twelve leukemia gene	6q23	Yes		B-ALL			Dom	T
SUFU	Suppressor of fused homolog (*Drosophila*)	10q24.32	Yes	Yes	Medulloblastoma	Medullo-blastoma	Medulloblastoma predisposition	Rec	D, F, S
SUZ12	Suppressor of zeste 12 homolog (*Drosophila*)	17q11.2	Yes		Endometrial stromal tumours			Dom	T
SYK	Spleen tyrosine kinase	9q22	Yes		MDS, peripheral T-cell lymphoma			Dom	T
TAF15	TAF15 RNA polymerase II, TATA box-binding protein (TBP)-associated factor, 68 kDa	17q11.1-q11.2	Yes		Extraskeletal myxoid chondrosarcomas, ALL			Dom	T
TAL1	T-cell acute lymphocytic leukemia 1 (SCL)	1p32	Yes		Lymphoblastic leukemia/biphasic			Dom	T
TAL2	T-cell acute lymphocytic leukemia 2	9q31	Yes		T-ALL			Dom	T
TCEA1	Transcription elongation factor A (SII), 1	8q11.2	Yes		Salivary adenoma			Dom	T
TCF1	Transcription factor 1, hepatic (HNF1)	12q24.2	Yes	Yes	Hepatic adenoma, hepatocellular ca	Hepatic adenoma, hepatocellular ca	Familial hepatic adenoma	Rec	Mis, F
TCF12	Transcription factor 12 (HTF4, helix-loop-helix transcription factors 4)	15q21	Yes		Extraskeletal myxoid chondrosarcoma			Dom	T

TCF3	Transcription factor 3 (E2A immunoglobulin enhancer-binding factors E12/E47)	19p13.3	Yes	pre-B-ALL	Dom	T
TCL1A	T-cell leukemia/ lymphoma 1A	14q32.1	Yes	T-CLL	Dom	T
TCL6	T-cell leukemia/ lymphoma 6	14q32.1	Yes	T-ALL	Dom	T
TEC	tec protein tyrosine kinase	4p12	Yes	Extraskeletal myxoid chondrosarcoma	Dom	T
TFE3	Transcription factor binding to IGHM enhancer 3	Xp11.22	Yes	Papillary renal, alveolar soft part sarcoma	Dom	T
TFEB	Transcription factor EB	6p21	Yes	Renal (childhood epithelioid)	Dom	T
TFG	TRK-fused gene	3q11-q12	Yes	Papillary thyroid, ALCL	Dom	T
TFPT	TCF3 (E2A) fusion partner (in childhood Leukemia)	19q13	Yes	pre-B-ALL	Dom	T
TFRC	Transferrin receptor (p90, CD71)	3q29	Yes	NHL	Dom	T
THRAP3	Thyroid hormone receptor-associated protein 3 (TRAP150)	1p34.3	Yes	Aneurysmal bone cysts	Dom	T
TIF1	Transcriptional intermediary factor 1 (PTC6,TIF1A)	7q32-q34	Yes	APL	Dom	T
TLX1	T-cell leukemia, homeobox 1 (HOX11)	10q24	Yes	T-ALL	Dom	T

(continued)

Symbol	Name	Location	Mutations		Tumor types (somatic mutations)*	Tumor types (germline mutations)*	Cancer syndrome	Mode	Mutation type(s)*
			Somatic	Germline					
TLX3	T-cell leukemia, homeobox 3 (HOX11L2)	5q35.1	Yes		T-ALL			Dom	T
TMPRSS2	Transmembrane protease, serine 2	21q22.3	Yes		Prostate			Dom	T
TNFRSF17	Tumor necrosis factor receptor superfamily, member 17	16p13.1	Yes		Intestinal T-cell lymphoma			Dom	T
TNFRSF6	Tumor necrosis factor receptor superfamily, member 6 (FAS)	10q24.1	Yes		TGCT, nasal NK/T lymphoma, skin squamous cell ca-burn scar-related			Rec	Mis
TOP1	Topoisomerase (DNA) I	20q12-q13.1	Yes		AML*			Dom	T
TP53	Tumor protein p53	17p13	Yes	Yes	Breast, colorectal, lung, sarcoma, adrenocortical, glioma, multiple other tumour types	Breast, sarcoma, adrenocortical carcinoma, glioma, multiple other tumour types	Li–Fraumeni syndrome	Rec	Mis, N, F
TPM3	Tropomyosin 3	1q22-q23	Yes		Papillary thyroid, ALCL			Dom	T
TPM4	Tropomyosin 4	19p13.1	Yes		ALCL			Dom	T

TPR	Translocated promoter region	1q25	Yes		Papillary thyroid			Dom	T
TRA@	T-cell receptor alpha locus	14q11.2	Yes		T-ALL			Dom	T
TRB@	T-cell receptor beta locus	7q35	Yes		T-ALL			Dom	T
TRD@	T-cell receptor delta locus	14q11	Yes		T-cell leukemia			Dom	T
TRIM33	Ttripartite motif-containing 33 (PTC7,TIF1G)	1p13	Yes		Papillary thyroid			Dom	T
TRIP11	Thyroid hormone receptor interactor 11	14q31-q32	Yes		AML			Dom	T
TSC1	Tuberous sclerosis 1 gene	9q34		Yes		Hamartoma, renal cell	Tuberous sclerosis 1	Rec	D, Mis, N, F, S
TSC2	Tuberous sclerosis 2 gene	16p13.3		Yes		Hamartoma, renal cell	Tuberous sclerosis 2	Rec	D, Mis, N, F, S
TSHR	Thyroid-stimulating hormone receptor	14q31	Yes	Yes	Toxic thyroid adenoma	Thyroid adenoma		Dom	Mis
TTL	Tubulin tyrosine ligase	2q13	Yes		ALL			Dom	T
USP6	Ubiquitin-specific peptidase 6 (Tre-2 oncogene)	17p13	Yes		Aneurysmal bone cysts			Dom	T
VHL	von Hippel–Lindau syndrome gene	3p25	Yes	Yes	Renal, hemangioma, pheochromocytoma	Renal, hemangioma, pheochromocytoma	von Hippel–Lindau syndrome	Rec	D, Mis, N, F, S
WAS	Wiskott–Aldrich syndrome	Xp11.23-p11.22				Lymphoma	Wiskott–Aldrich syndrome	X-linked recessive	Mis, N, F, S
WHSC1	Wolf–Hirschhorn syndrome candidate 1(MMSET)	4p16.3	Yes		MM			Dom	T

(continued)

Symbol	Name	Location	Mutations		Tumor types (somatic mutations)*	Tumor types (germline mutations)*	Cancer syndrome	Mode	Mutation type(s)*
			Somatic	Germline					
WHSC1L1	Wolf–Hirschhorn syndrome candidate 1-like 1 (NSD3)	8p12	Yes		AML			Dom	T
WRN	Werner syndrome (RECQL2)	8p12-p11.2		Yes		Osteosarcoma, meningioma, others	Werner syndrome	Rec	Mis, N, F, S
WT1	Wilms tumour 1 gene	11p13	Yes	Yes	Wilms, desmoplastic small round cell tumor	Wilms	Denys–Drash syndrome, Frasier syndrome, Familial Wilms tumor	Rec	D, Mis, N, F, S
WTX	Family with sequence similarity 123B (FAM123B)	Xq11.1	Yes		Wilms tumour			Rec	F, D, N, Mis
XPA	Xeroderma pigmentosum, complementation group A	9q22.3		Yes		Skin basal cell, skin squamous cell, melanoma	Xeroderma pigmentosum (A)	Rec	Mis, N, F, S
XPB	Excision repair cross-complementing rodent repair deficiency, complementation group 3 (xeroderma pigmentosum group B complementing)	2q21		Yes		Skin basal cell, skin squamous cell, melanoma	Xeroderma pigmentosum (B)	Rec	Mis, S

XPC	Xeroderma pigmentosum, complementation group C	3p25		Yes	Skin basal cell, skin squamous cell, melanoma	Xeroderma pigmentosum (C)	Rec	Mis, N, F, S
XPD	Excision repair cross-complementing rodent repair deficiency, complementation group 2 (xeroderma pigmentosum D)	19q13.2-q13.3		Yes	Skin basal cell, skin squamous cell, melanoma	Xeroderma pigmentosum (D)	Rec	Mis, N, F, S
XPG	Excision repair cross-complementing rodent repair deficiency, complementation group 5 (xeroderma pigmentosum, complementation group G (cockayne syndrome))	13q33		Yes	Skin basal cell, skin squamous cell, melanoma	Xeroderma pigmentosum (G)	Rec	Mis, N, F
ZNF145	Zinc finger protein 145 (PLZF)	11q23.1	Yes	APL			Dom	T
ZNF198	Zinc finger protein 198	13q11-q12	Yes	MPD, NHL			Dom	T
ZNF278	Zinc finger protein 278 (ZSG)	22q12-q14	Yes	Ewings sarcoma			Dom	T
ZNF331	Zinc finger protein 331	19q13.3-q13.4	Yes	Follicular thyroid adenoma			Dom	T
ZNF384	Zinc finger protein 384 (CIZ/NMP4)	12p13	Yes	ALL			Dom	T

(continued)

Symbol	Name	Location	Mutations		Tumor types (somatic mutations)*	Tumor types (germline mutations)*	Cancer syndrome	Mode	Mutation type(s)*
			Somatic	Germline					
ZNF9	Zinc finger protein 9 (a cellular retroviral nucleic acid-binding protein)	3q21	Yes		Aneurysmal bone cysts			Dom	T
ZNFN1A1	Zinc finger protein, subfamily 1A, 1 (Ikaros)	7p12	Yes		ALL, DLBL			Dom	T

A Amplification; AEL Acute eosinophilic leukemia; AL Acute leukemia
ALCL Anaplastic large-cell lymphoma; ALL Acute lymphocytic leukemia; AML Acute myelogenous leukemia
AML* Acute myelogenous leukemia (primarily treatment associated); APL Acute promyelocytic leukemia; B-ALL B-cell acute lymphocytic leukemia
B-CLL B-cell Lymphocytic leukemia; B-NHL B-cell non-Hodgkin lymphoma; CLL Chronic lymphatic leukemia
CML Chronic myeloid leukemia; CMML Chronic myelomonocytic leukemia; CNS Central nervous system
D Large deletion; DFSP Dermatofibrosarcoma protuberans; DLBL Diffuse large B-cell lymphoma
DLCL Diffuse large-cell lymphoma; Dom Dominant; E Epithelial
F Frameshift; GIST Gastrointestinal stromal tumour; JMML Juvenile myelomonocytic leukemia
L Leukaemia/lymphoma; M Mesenchymal; MALT Mucosa-associated lymphoid tissue lymphoma
MDS Myelodysplastic syndrome; Mis Missense; MLCLS Mediastinal large-cell lymphoma with sclerosis
MM Multiple myeloma; MPD Myeloproliferative disorder; N Nonsense
NHL Non-Hodgkin lymphoma; NK/T Natural killer T-cell; NSCLC Non-small-cell lung cancer
O Other; PMBL Primary mediastinal B-cell lymphoma; pre-B-All Pre-B-cell acute lymphoblastic leukemia
Rec Reccesive; S Splice site; T Translocation
T-ALL T-cell acute lymphoblastic leukemia; T-CLL T-cell chronic lymphocytic leukemia; TGCT Testicular germ cell tumour
T-PLL T-cell prolymphocytic leukaemia

Index

A
Aaronson, Stuart 56
Aberrant crypt focus (ACF)
  definition of 35
  RAS mutations in 75
Activating mutations 49
Activation loop 184
Acute lymphocytic leukemia (ALL) 244
Acute myelogenous leukemia, in FA 156
Adenomas
  colorectal 43
  tumor suppressor gene inactivation in 120
Aflatoxin B1 (AFB1) 23, 255
AKT 191
  in endometrial cancer 236
Alu repeats 6
Aneuploidy,
  definition of 126
  p53 loss and 131
  possible mechanisms of 133–135
Angiogenesis 211
APC
  and colorectal tumorigenesis 44
  as a biomarker 267
  mutations in 89
  positional cloning of 88, 89
  pre-mutations of 91
  WNT signaling and 194
Apoptosis 28
  downregulation by AKT 191
  loss of sensitivity to 174
  p53 induction of 211
  pathways to 214–218
  therapeutic targeting of 276
Apoptosome 216
Asbestos 24
Ascertainment bias 95
Ashkenazi Jewish population
  BS in 165
  founder effects 104
  founder FA mutations in 159
  germline APC mutations in 91
Astrocytoma 253
Ataxia Telangiectasia-like disorder (ATLD) 163
Ataxia-telangiectasia (AT) 160
Ataxia-telangiectasia mutated (ATM) 245
  activation of p53 by 206
  cloning of 162
  in breast cancer 234
Ataxia-telangiectasia and Rad3 Related (ATR) 208
Atypical nevi 105
AXIN 195

B
Bang, Olaf 50
Bannayan-Riley-Ruvalcaba syndrome 110
Barbacid, Mariano 56
Barrett's esophagus 264
BAX 217
BCL2 217
BCR-ABL 64, 185, 269
Beach, David 104
Benedict, William 83
Benign prostatic hyperplasia (BPH) 231
Benzo[a]pyrene diol epoxide (BPDE)
  lung cancer and 230
  mutagenesis and 20
Biomarker 262
Bishop, J. Michael 53
BLM 214
Bloom syndrome (BS) 163
Bootsma, Dirk 149
Boveri, Theodor 126
BRAF 189
  in melanoma 242
  in thyroid cancer 251

BRCA1 233, 234
discovery of 102
DNA repair and 214
prostate cancer and 233
BRCA1 and *BRCA2*, genetic testing 261
BRCA2 234
discovery of 102
DNA repair and 214
Germline mutations in FA 157
Breast cancer 102, 233–235
Li Fraumeni syndrome and 102
male 103
Burkitt lymphoma 238

C
c.711+4A>T mutation 157
Cadherins 196
Cancer genes
acquisition of 4
definition of 4
versus benign genetic variants 10
Cancer stem cells 40
Candidate gene approach 34
Caretakers, definition of 123
Carrier identification 260
Carson, Dennis 104
Cascade 178, 216
Caspases 216
β-catenin (*CTNNB1*) 195, 216
in endometrial cancer 236
in ovarian cancer 243
in thyroid cancer 251
Cavanee, Webster 85
Cdc25 phosphatases 221
CDH1 253
CDK4 241
CDKN1A (p21)
activation by TGF-β 200
cell cycle regulation by 223
inhibition by AKT 191
CDKN2A (p16)
cell cycle regulation by 222
in oropharyngeal cancer 248
in bladder cancer 239
in GBM 254
in melanoma 240
in pancreatic cancer 247
mouse models of 118
CDKN2B (p15), cell cycle regulation by 222
Cell cycle arrest 210
Cetuximab 275
Checkpoint kinase 1 (Chk1) 208
Checkpoint kinase 2 (Chk2) 208
Checkpoints 223
Chemotherapy 263
Chromosomal instability (CIN), definition of 129
Chronic inflammation 24
in liver cancer 256
role in lymphoma 238
role in pancreatic cancer 247
Chronic lymphocytic leukemia (CLL) 244
Chronic myelogenous leukemia (CML) 245, 269
translocations in 63
C-KIT 271
Clastogen 22
genetic instability caused by 156
Clonal evolution 27
Clonal nature of cancer, evidence for 28
Clonal selection 29
C-MYC 57-60
function of 201
in breast cancer 235
in leukemia 245
in medulloblastomas 255
stabilization of 202
transcriptional activation of 202
Cockayne syndrome 153
Collins, Francis 115
Colorectal cancer 35–39
aneuploidy in 131
inactivation of WNT signaling in 196, 197
oncogenes and 74
tumor suppressor mutations and 120
Complex atypical hyperplasia (CAH) 235
Compound heterozygosity
in ATM 162
in FA 159
Congenital hypertrophy of the retinal pigment epithelium (CHRPE) 89
Cooper, Geoffrey 55. 56
Cowden disease 109, 112
Cowden syndrome 252
CpG islands 18
Croce, Carlo 217
Crosstalk 178
Cryptic splice sites 9
CTNNB1 (see β-catenin)
Cyclin D (CCND)
control of cell cycle by 220
in breast cancer 235
in Mantle cell lymphoma 238
in oropharyngeal cancer 248
inhibition by AKT 191
Cyclin dependent kinases 218
Cyclin E (CCNE) 220
Cyclins 218

Cyctochrome C 216
Cytogenetic abnormalities, definition 5
Cytokines 173, 198

D
De la Chappelle, Albert 140
Deamination 15
Death-inducing signaling complex (DISC) 216
Deletion 5
DeSanctis-Cacchione syndrome 153
Diagnosis 259
Disheveled 195
DNA damage signaling network 207, 268
DNA methylation
  gene silencing and 18
  transitions and 15
DNA repair
  DNA damage signaling and 214
  p53 induction of 211
DNA replication error rate 136
Double minutes 58
Dryja, Thaddeus 85
Ductal carcinoma in situ (DCIS) 233
Duesberg, Peter 135

E
E2F 219
Early detection 259
Ectodomains 275
Epidermal growth factor receptor (EGFR)
  activation of 183
  activation of PI3K by 190
  activation of RAS by 187
  inhibition by cetuximab 275
  inhibition by gefitinib 273
Ehrlich, Paul 259
Ellerman, Willhelm 50
Endometrial intraepithelial carcinoma (EIC) 235
Eng, Charis 110
Epigenetics, definition 17
ERBB2
  activation of 183
  discovery of 60
  amplification of 72
  in breast cancer 235
  inhibition by trastuzumab 275
Erikson, Ray 180
Esterase D (ESD), linkage to RB 83
Ewing's sarcoma 67, 69
EWS-FLI1 68, 72
Excision repair cross complementing genes (ERCC) 149
Exon skipping 8, 13
Extracellular signal regulated kinases (ERK) 188
Extrinsic pathway 215, 276

F
FA core complex 214
Familial adenomatous polyposis (FAP) 43-44
  mouse models of 119
Familial atypical multiple mole syndrome (FAMMM) 240, 247
Familial medullary thyroid cancer (FMTC) 73
FANCB 159
Fanconi anemia (FA) 156, 245
  DNA repair and 214
Fanconi, Guido 156
FDXR 218
Follicular lymphoma 237
FOXO transcription factors 192
Frameshift mutation, definition 13
Fraumeni, Joseph 94
Frizzled 195
Fung, Yuen-Kai 85

G
Gardner's syndrome 89
Gastrointestinal stromal tumor (GIST) 271
Gatekeepers, definition of 123
Gatti, Richard 162
GDP/GTP cycle 186
Gefitinib 273
Gene amplication, oncogene activation and 58
Gene conversion 159
Genetic testing 259
Germ cells 4
German, James 166
Germline mutations
  cancer risk and 41, 42
  definition of 4
Glioblastoma multiforme (GBM) 253
Global genome repair (GGR) 151
Glucose metabolism 29, 30
  p53 regulation of 211
Glycolysis 29
GSK3 kinase 195
GTPase activating protein (GAP) 187
GTPase 186
Guanine nucleotide exchange factor (GEF) 186

H
Hamartoma 111
Hanaoka, Fumio 150
Hansenmann, David 126
Haploid genome size 109
Harris, Henry 77
HBX 255
Helicobacter pylori 252
Hepatitis viruses 255
HER2/neu (see ERBB2)
Hereditary diffuse gastric cancer (HDGC) 253
Hereditary nonpolyposis colorectal cancer (HNPCC) 139–146
  bladder cancer in 239
  clinical features of 43–45
  endometrial cancer 235
  pancreatic cancer 247
  prevalence of 143
  stomach cancer 253
Hereditary pancreatitis 247
hMLH1 141
hMSH2 141
hMSH6 141
Hodgkin lymphoma 237
Hoeijmakers, Jan 149
Homologous recombination
  DNA repair and 22
  Bloom syndrome and 163–164
hPMS2 141
H-RAS
  in bladder cancer 239
  in oropharyngeal cancer 248
Human genome, size 10
Human papilloma virus 94, 248
Hunter, Tony 180
Hypoxia 30

I
Icelandic population genetics 11, 103
Imatinib 269
In vitro transformation 56
Incidence, definition of 229
Insertion 5
Intrinsic pathway 215
Ionizing radiation
  mutagenesis and 20
  as therapy 268

J
Juvenile polyposis syndrome (JPS) 111

K
Kaposi, Moriz 147
Kastan, Michael 206
Kern, Scott 111
Kinase, definition of 180
King, Mary-Claire 102
Kinzler, Kenneth 38, 128
Knockout mice 118
Knudson, Alfred 80
Kolodner, Richard 141
K-RAS 54–57
  as a biomarker 266
  in endometrial cancer 236
  in oropharyngeal cancer 248
  in ovarian cancer 242
  in pancreatic cancer 245
  mutations in cancer 189

L
Landscapers, definition of 123
Lane, David 91
Lee, Wen-Hwa 85
Legerski, Randy 149
Lengauer, Christoph 128
Levine, Arnold 91
Li Fraumeni syndrome (LFS) 94–98
  mouse models of 118
  leukemias in 245
  GBM in 255
Li, Da-Ming 109
Li, Frederick 94
Linn, Stuart 150
Liporeceptor-related protein (LRP) 195
Lobular carcinomas in situ (LCIS) 233
Loss of heterozygosity (LOH)
  CIN and 132
  mechanisms of 86, 87
Low fidelity DNA repair, mutagenesis and 17
Lymphomas, chromosomal translocations in 63
Lynch syndrome (see hereditary nonpolyposis colorectal cancer)
Lynch, Henry 139

M
Magic bullet 259
Malignancy, definition of 3
MALT 237
Mantle cell lymphoma 238

MAPKK proteins (see MEK proteins)
Marx, Stephen 116
MDM2 feedback loop 211
MDM2 94, 97, 205
  in Hodgkin lymphoma 238
  SNP309 in 97
Medulloblastomas 255
MEK proteins (MAPKK proteins) 187
Melanocyte 240
Melanoma 104, 106
  in XP patients 147
MEN1, cloning of 116
Meningioma 255
Mesothelioma 24
MET
  cancer predisposition and 73
  in kidney cancer 244
Microdeletion 5
Micro-insertion 5
Microsatellite instability (MSI) 140
Mismatch repair (MMR) 138
Missense mutation, definition 13
Mitogen-activated protein kinases (MAP) 188
Mitotic recombination 86
MLL 245
Monogenic diseases 3
Mosaic variegated aneuploidy 133
Mouse models 118
MRN complex 206, 214
mTOR pathway 192
Multiple endocrine neoplasia type 1 (MEN 1) 116
Multiple endocrine neoplasia type 2 (MEN2) 73, 251
Mutagens 18
Mutations
  driver and passenger 32
  number of in cancer 31
  types of 5-7
MYC genes, discovery of 57
MYC-associated protein (MAX) 201

N
Nakamura, Yusuke 89
Neisser, Albert 147
Neoplasm, definition of 2
Neuroblastoma 59
Neurofibromatosis 1 (van Recklinghausen neurofibromatosis) 144, 255
Neurofibromatosis 2 115, 255
Neuroglia 253
NF1,
  cloning of 114
  function of 189
Nijmegen breakage syndrome (NBS) 163
Nitric oxide 173
N-MYC
  amplification 59
  in medulloblastomas 255
Non-Hodgkin lymphoma 237
Non-homologous end joining
  ATM and 207
  DNA repair and 22
Non-protein coding genes 9, 10
Nonsense-mediated RNA decay 14
Nonsynonymous mutations 71
Nucleotide biosynthesis 16
Nucleotide excision repair (NER) 138
Null allele, definition 15

O
Odds ratio, definition of 100
Okada, Yoshio 150
Oncogene, definition of 49
Open reading frame, definition 7
Oxidative phosphorylation 29

P
p14 alternative reading frame (p14 ARF) 107
p15 (*CDKN2B*) 107
p16 (CDKN2A) 104
  P53 mutations and 22, 23
p53
  activation of 204–209
  as a biomarker 264
  discovery and cloning 91, 92
  genes induced by 209
  in Barrett's esophagus 264
  in breast cancer 234
  in endometrial cancer 236
  in gastric carcinomal 252
  in GBM 254
  in liver cancer 255
  in lung cancer 230
  in oropharyngeal cancer 247
  in ovarian cancer 242
  mutations in 92, 93
  regulation of the cell cycle by 222
  repression by MDM2 212, 213
Pancreatic intraepithelial neoplasia (PanINs) 245
Parsons, Ramon 109

Penetrance 4, 100
Perucho, Manuel 140
Philadelphia chromosome 63, 245
Phosphatase, definition of 180
Phosphatidylinositol 3-kinases (PI3K), functions of 190
PIK3CA
activation of 193
in breast cancer 235
in colorectal tumorigenesis 75
in endometrial cancer 236
in GBM 254
in lung cancer 231
mutations of 71
PIP_2/PIP_3 190
Pleckstrin homology domain 191
Pleiotropy, XP genes and 154
Poikilodermia 147
Point mutation, definition 15
Polymorphisms, definition 10, 11
Polyps, colorectal 43
Posttranslational modification 175
Prakash, Louise 150
Prevalence, definition of 229
PRL-3 75
Progeroid syndromes 166
Prognosis 259
Prostatic intraepithelial neoplasia (PIN) 231
Protein half-life 176
Protein phosphorylation 180
Protein structure 13
Protein tyrosine kinases 181
Proto-oncogenes
activation of 71
definition of 49
PRSS1 247
PTEN
discovery of 108
in endometrial cancer 236
in GBM 253
in ovarian cancer 242
inactivation of 193
PUMA 217

R
Radiomimetics 268
Radioresistant DNA synthesis (RDS) 161
RAF kinases 187
Random aneuploidy 135
RAS genes
alterations in cancer 189
discovery of 54–57
functions of 186
in colorectal tumorigenesis 75
Rational therapies 259
RECQ helicase 166
Recurrence 264
Relative risk, definition of 100
Repetitive elements 6
Representational difference analysis 109
Restriction fragment length polymorphism (RFLP) 85
RET 251
cancer predisposition and 73
Retinoblastoma, features of 79
Retinoblastoma gene (*RB*)
cloning of 80–86
in lung cancer 230
in melanoma 240
regulation of the cell cycle by 218
Rothmund-Thomson syndrome 170
Rous sarcoma virus 50, 185
Rous, Peyton 50

S
Schellenberg, Gerard 168
SCO2 211
Segmental progeroid syndromes 168
SFN (14-3-3σ) 224
Shiloh, Yosef 162
Silent mutations, definition 12
Single nucleotide polymorphisms (SNPs) 10
Single nucleotide substitutions, definition 15
Sister chromatid exchange 164
Skolnick, Mark 102, 104
Slipped mispairing model 16
SMAD4 (*DPC4*) 111, 199
in pancreatic cancer 247
proteins 199
Somatic cells 4
Somatic mutation, definition of 4
SOS proteins 186
Splice acceptor mutations 8
Splice donor mutations 8
Sporadic cancers, definition 43
SRC homology domains (SH) 184
SRC, discovery of 53
Staging 263
Stalled replication forks, misincorporation and 17

Steck, Peter 109
Sun, Hong 109
Synonymous mutations 71

T
TCF4 197
TCL1 245
Telomeres 168
Tetraploidy, origin of 135
TGF-β signaling 198
Thibodeau, Stephen 140
Thompson, Larry 149
Tissue homeostasis, 28
Tobacco smoke, *P53* mutations and 19
TRAIL 215, 276
Transcription Factor II H subunits 152
Transition zone
  in uterine cervix 248
  in prostate 231
Translocation, definition of 62
Trascription coupled repair (TCR) 151
Trastuzumab 275
Trichothiodystrophy 153
TSC1, *TSC2* 192
Tumor necrosis factor 215
Tumor suppressor gene, definition of 77
Tumorigenesis, definition of 2
Tumorigenicity, definition of 78
Turcot syndrome 142, 255
Two-hit hypothesis 79, 80

U
Ulcerative colitis
  in colorectal cancer 24
  tetraploidin 136
Ultraviolet light (UV),
  P53 mutations and 20
  Photoproducts 20
  Signature mutations 20
Unscheduled DNA synthesis (UDS) 148
UV photoproducts, removal by NER
    system 146

V
Varmus, Harold 53
Viral infection 4, 52
Viral oncogenes 52
Vogelgram 38
Vogelstein, Bert 38, 89, 128, 140, 141
Von Hippel-Lindau renal carcinoma (VHL) 243

W
Warburg effect 29, 211
Warburg, Otto 29
Warthin, Aldred 139
Weinberg, Robert 55, 85
Werner syndrome 167, 252
White, Ray 85, 89, 115
Wigler, Michael 56, 109
WNTs 194
Wobble position 12
WRN 168, 169

X
Xeroderma pigmentosum (XP) 146

Z
Zollinger-Ellison syndrome 116